

; CURRENT APPLICATION NUMBER: US/10/463,957
; CURRENT FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: 60/389,685
; PRIOR FILING DATE: 2002-06-17
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 5
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-10-463-957-5

Query Match 20.0%; Score 1734; DB 12; Length 420;
Best Local Similarity 78.3%; Pred. No. 1.7e-100;
Matches 329; Conservative 48; Mismatches 43; Indels 0; Gaps 0;

QY 1 MNLVNRKQLEKMANVRFRQTQDEYVAILDALAEYHNHNSSENTVVEKYLKLDINSLTDIYI 60
DB 1 MSLVNRKQLEKMANVRFRQTQDEYVAILDALAEYHNHNSSENTVVEKYLKLDINSLTDIYI 60

QY 61 NTYKSGRNKALKKFKFYEYLTMEVLELKNNSLTPVEKNLHFWIGQINDTAINYNQWKD 120
DB 61 DTYKSGRNKALKKFKFYEYLTMEVLELKNNSLTPVEKNLHFWIGQINDTAINYNQWKD 120

QY 121 VNSDYTVKVEYDGNALFINTLKKTIVESATNTTLESFRENLDNPEFDYK 180
DB 121 VNSDYTVKVEYDGNALFINTLKKTIVESATNTTLESFRENLDNPEFDYK 180

QY 121 VNSDYNVNFYDGNALFINTLKKTIVESATNTTLESFRENLDNPEFDYK 180
DB 121 VNSDYNVNFYDGNALFINTLKKTIVESATNTTLESFRENLDNPEFDYK 180

QY 181 DKQGHFDYKSGIENPEFIIDIIKTLYLSNEYSKDLALANKYIEESLNKITTANGNDI 240
DB 181 DKQGHFDYKSGIENPEFIIDIIKTLYLSNEYSKDLALANKYIEESLNKITTANGNDI 240

QY 241 RNLEKFADEDLVLYNGELVERNLAAASDILRISMLKEDGGVYLDVILPGIOPDLFKS 300
DB 241 RNLEKFADEDLVLYNGELVERNLAAASDILRISMLKEDGGVYLDVILPGIOPDLFKS 300

QY 301 INKPDSTINTSWEMKLEAMTKYKEIPIGYTSKNFMDLDEEVSQSPESALSSKSKSEIF 360
DB 301 INKPDSTINTSWEMKLEAMTKYKEIPIGYTSKNFMDLDEEVSQSPESALSSKSKSEIF 360

QY 361 LPDIDIKVPLEVKIAFANNSVINOALISLSDVCSDLVINOIKRYKIINDNLNPSINE 420
DB 361 LPDIDIKVPLEVKIAFANNSVINOALISLSDVCSDLVINOIKRYKIINDNLNPSINE 420

; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637
; CURRENT APPLICATION NUMBER: US/10/463,957
; CURRENT FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: 60/389,685
; PRIOR FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-10-463-957-7

Query Match 8.9%; Score 769; DB 12; Length 170;
Best Local Similarity 85.3%; Pred. No. 1.5e-40;
Matches 145; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

QY 1 MNLVNRKQLEKMANVRFRQTQDEYVAILDALAEYHNHNSSENTVVEKYLKLDINSLTDIYI 60
DB 1 MSLVNRKQLEKMANVRFRQTQDEYVAILDALAEYHNHNSSENTVVEKYLKLDINSLTDIYI 60

QY 61 NTYKSGRNKALKKFKFYEYLTMEVLELKNNSLTPVEKNLHFWIGQINDTAINYNQWKD 120
DB 61 DTYKSGRNKALKKFKFYEYLTMEVLELKNNSLTPVEKNLHFWIGQINDTAINYNQWKD 120

QY 121 VNSDYTVKVEYDGNALFINTLKKTIVESATNTTLESFRENLDNPEFDYK 170
DB 121 VNSDYNVNFYDGNALFINTLKKTIVESATNTTLESFRENLDNPEFDYK 170

RESULT 14
US-10-114-170-257
; Sequence 257, Application US/10114170
; Publication No. US20030023075A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Buriand, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/114,170
; FILING DATE: 01-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/453,702
; FILING DATE: 03-DEC-1999
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 257:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3169 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 257:
US-10-114-170-257

Query Match 5.9%; Score 512.5; DB 14; Length 3169;
Best Local Similarity 20.2%; Pred. No. 1.6e-22;
Matches 422; Conservative 333; Mismatches 720; Indels 619; Gaps 104;

QY 14 YVKFEI---QEDEYVAIINAL--EEYHNHNSSESVVEKYLKLDINNLDTNLTNTYKSGR 68
DB 160 YIKIRKTRGAEDQTITQSLIINELLNGVDNTI--PFQKISELNDIHSVENMIKNSR 218
QY 69 N--KALKKFKFYEYLTMEVLELKNNS-----LTPVEKNLHFWIGQ 106
DB 219 KGIEILVKQGLLSLLINDNKGKQISDNASKIINLLGIEYQSHKVDIEPFIHAVVAGA 278
QY 107 INDNTAINYNQWKDVSNTYTVKFDVDSNAF-----LINTLKT----- 144

Db 279 PPONTREYITAFNTYKDYTYLLWIDPNAFAAKFGSILKNIAVMYIMRLRTPHPLAE 338
Qy 145 -----IVESATNNTLE--SFRNLNDPEFY-----NKPYKRMELIYDQKHFI 187
Db 339 EMNEVILKIQNIQNETIEFKETRELERKLEENRYKSLTSETKEFNFFLESIMGMQDNYP 398
Qy 188 DYKSOIEENPEFI-----DNIUK--TYLSNEYSKDLEALNKYIEESLNKITTANNGN-- 238
Db 399 TYCISNGISNTDOLSRUDFTNVLKUSPEVONDFKSTVEKVKRDIILLKNTISQKEDRF 458
Qy 239 ---DIRNLEKFADEBLVRLYNQELVERNLAASDIRLSMLKEDGGVYLDVVDILPGIQP 295
Db 459 QLRDINTLESFKKPDQFFVQOEMLEWNAASDQVRINILKEYGGIYTDITDILPAYSD 518
Qy 296 DLFSKNKPSITNTSWMEMIKLEAIMKY--EYIPG--YTSKNPMDLDEEVQBSFSAIS 351
Db 519 KVSQIINE-XSDSKRFFEDLKRLIISSESILSIKGEKYSIKG-DGLDETTLNQLNNIL- 575
Qy 352 SKSDKSEIFPLDLDIKVSPLEVKIAPANNVI-----NQALLSLKSDVC 395
Db 576 SEIEK-----LTDIDY-FKPVETKVVRDTFKIPRYQKWENTWNIRGNFMTLHGSKC 630
Qy 396 SDLVINOIKRYKILNDNLNPSINEGTDFTNTMKIFSDKLASISNE-----DNM 444
Db 631 IDFILSGKKQYLEL-QRIRDNISYNNLFYTT-----EDLKSNNVAIGGIPAKKYLEHG 684
Qy 445 MFMIKITNLYKVPAPDVRSTINSGP----- 471
Db 685 LF-----SEYRQDGTIPYVSTLNTSGPMIMROMKMYKYSIGRICEVHDKNKLSDVNFL 740
Qy 472 GYVYGAQDLMF---KDNSTNIHLBPE-----LNFEPFKTKISQLT-----EQEITS 518
Db 741 GYVASSKNKNSFWNLNPSVSGINDITPDDESSWAVRNNDINKILFEKINCHVPEKLUPT 800
Qy 519 LWSFNQARAKSQFEYKKGYPEGALGEDDNDPAQNTVLDKDYVSKILSMLKTRKEYI 578
Db 801 LY-----YELDSRFFQGW-----DNKSIKHVTEINKDLI--KOINLLITSSNIDV 844
Qy 579 HYIVOLOQDKISYEASCNLFKDPYSILYOKNIEGSETAYVYVADAIEKIDKXIPY 638
Db 845 KLLIKL--DRELYAISXSI-----DNPLALRSIRTLQQLANYVTSNTEPENTINFIY 896
Qy 639 QISNKN-----IKLTFIGHGKBENFTDTTANLDVDS-----LSSEIET 677
Db 897 DYPKQKQDLLASAUKL-----FSRNDADTKIIWYNSYMEKNVFLREVIS 942
Qy 678 ILNLAKADLSPYIEINLLGNMFSYSISABETYPGKLLKIKDRVSELMPSISQDSI-- 735
Db 943 VLRSKKVD---SYINEN-----KKNLSKEDA--GALRDYAKUKMKELFSLMLDDGGYK 990
Qy 736 TVSANQYEVRIINEGKREILDHSGKWINKESIIKDISKEYISFNP-----K 783
Db 991 IITWAY---IKERDKL-----SGIYNIENSIISCHESFIIIRSNQHEWGDLSVQFK 1042
Qy 784 ENKIIVKSKYLHELSTLQEIIRNANSDDILEKKVMUTECEINVAS-----NTDRQIVE 838
Db 1043 KFEFVYKSE-LSSAKSIPDDIKNK-YITDPETKRNVLVHQLDSDIKERIAFLDISHVAYP 1100
Qy 839 GRIBKXONTSDSINYINKNEFKLIESI-----SDSLYDLKHQNL 878
Db 1101 GSLEKQLQS-----GVYFSDINIIAYLLASYGVSGHSHYVVPAPSDKLELLRHTK 1156
Qy 879 DSHRISPEIDSKTENGPRIRINKETGNSI---FIETKEIFSYATHISKEISNIKDT 935
Db 1157 SNSEMI--EKITP-----YVVDILSNVSNLPPPLSEBQKILNDIKLEISKVS---EQ 1207
Qy 936 IFDNVNGKLVKVNLDAAHEVNTLSAFTIQLIBYNTTK-----ESLGNLSV----- 983
Db 1208 YPKLUEQKSSVIGIKYVDFRYNENFLSLPINQNLTLPMRYFEMULDIHIGIEN 1267
Qy 984 -AMKVOVYALPSTGTNTTDSKVELV-----STAL-DETIDL 1021

Db 1368 KANREFIYKFSSESLNLDLINDERVNLGLIKKYKVLSELSEIHRTLTNTSTSPADISIPL 1327
Qy 1022 LPTLSEGLPIIATII-----DGVSLGAIAK-----ELSETNDPLLR 1057
Db 1328 LQTI---CPSITTIKKTEYVGHQLTNMTVSVVXPYDFSNLGAINSIDKSVSDVPALH 1384
Qy 1058 QEIE-AKIGIMAVNLTAASTAIVTSAL-----GIASGSILL 1093
Db 1385 TIVEQAKYNLLSMWDFYNTHASIWDTIARQKSTNIEFHPQSLLFDRDSKGKCLGSLLY 1444
Qy 1094 VPLAGISAGIPSLANN---ELILODKATKVIDYFKHISLAETEGATFLDDDKIIMPQDD 1149
Db 1445 LDTGGYGGYQKLRHNDITASTLYQTK-----YNDNLKLSNRDDFLRKTQRIITWSNE 1498
Qy 1150 LVLSEIDNNNSITLKGCEIWRAGEGSGHTLTDI--DHFFSPSSITTYRKPWLSIYDVNL 1207
Db 1499 L-----GNRELKNAQLEVLELX--DPILTEGLYQRRISLITTEYHSLAQISSL 1548
Qy 1208 IKK-----EKIDFSKOLMVLNANRVFGYEMGWTPGFRSLDNDGTKLLD 1252
Db 1549 FWRVTDPNFGCHDFHSLAQALTFIKNI-----TSNR-----NPSLSYSGSI 1589
Qy 1253 RIRDHY-EGQFYWYFAFIADALITKLKPRYEDTNVRINLDG--NTRSEIYVPIVITTEQI 1308
Db 1590 -VKIYFSESLANWKYIKLPLVQTGSLRDIYLTPEKLSSTGGSINIMGHLPV 1642
Qy 1309 RKNLSYSP-YSGSGSYSLSPYNNIDNLNVENDTWIDVNVVXNNTIESDBIQ----- 1363
Db 1643 ---STIYDIGV-----INGNRISEST---DVKNKIRSLKINGDILQHYIN 1682
Qy 1364 ---KGLIENILSKINIEDNKI-----ILN--NHTINPY- 1392
Db 1683 THYLSEEQTKIKDIVDFLGIQDNTIKVLESIDKIPSEIQOPLHSILSROKHEVKNLLS 1742
Qy 1393 GOINE-SNRF-----ISLTFSEILEDINI-----IIEI---DLVSKSYKILLSGNCKML 1436
Db 1743 GULDBFSNKLKQGLSKLNTLVSNVNFKESKINSITVEVTDTLQGLRYRVIDIDTRVIGL 1802
Qy 1437 I--ENSSDIQKIDHIGFNGEHQKYPISYIDNETKNGFYDYSK---KEGLFTAEPFN 1490
Db 1803 TFKEGINSUSEALEHNN-----IDAMSVIGLVQYARMIKWMDNISAIIDAG 1849
Qy 1491 E-SIIRNIWMPDSNLIYSSKDLKIRIINKGDYK--LLIGNYFKDDMKVLSLFTIEDTN 1548
Db 1850 AVSDIKNIV-----DKFLGGILTLTNVYNVPGVSGASLEGFISGLEVCAS----- 1897
Qy 1549 TIKLNGV---YLDENGVAQILKPMNNAKSALNTSNLNFLESINIKNIFYNLD----- 1600
Db 1898 --RMGSGTAGRYL--SNVAKVIKL-----PLLDI--GINIWSLYDSSLNEAKAT 1939
Qy 1601 PNIEFI---LDTNF-IISGNSIGQFE---LICDKKNIOQYFIFNEKIKETSITLYV-- 1650
Db 1940 TQIETISTAIDVSFSSINTALSIGAIVPPLAIA-----IVPITI-FSHEVKNYAVYVNO 1993
Qy 1651 -GNRQNLIVEPSVHLDD-----SGNISSTVNFQKLYG---IDRYVNVKVII 1694
Db 1994 INERHKLWEAEKYLIDNGSAKVLKINKATGIIDLNNQVNLGYLDMRENPPIL 2047

RESULT 15
US-10-282-122A-76865
; Sequence 76865, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlser, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant

```

; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 78665
; LENGTH: 4688
; TYPE: PRT
; ORGANISM: Ureaplasma urealyticum
; US-10-282-122A-78665

Query March 5.08; Score 431.5; DB 12; Length 4688;
Best Local Similarity 19.18; Pred. No. 3.4e-17;
Matches 420; Conservative 329; Mismatches 743; Indels 709; Gaps 103;

QY 13 VYVFRQDEBYVAILNALE-EYHNMSESSVWEKYLKLDINLNTNYLT---YKSGR 68
D 403 LYEVIGAKTNEVDVNLKQIPHNRS-----INLNSALNTPQYTKNGD 449
QY 69 NKALKKFEKYLME---VLELKNS---LTPVEKNLHPFIWGGQINDTAIYNQWKD 120
D 450 INLIKAVPYVNVQVYGFQKQNNQHQILAKVKQGTAFDQALNNNSVSLDKIVS 509
QY 121 VNSDYTVKVDVSNAPLINTLKTIVSANTNLESFR-----ENL----- 161
D 510 VSNPQNVLV---SNFDLTSKQKQILKPAANASVDSKTKQILENLNLDLQKLVATFV 566
QY 162 --NDPEFDYKFKRVEIIVDKQ---KHFIDYKSKQIEN----- 197
D 567 DNNDKEYKVVANVDQNNKIIFSDNSDLPGVI-YHLAKVNNNDLKNVNLNDFELDKIID 625
QY 198 -----PEFTIDN---IKTYLSNEYSKOL--EALNKY-----IBESLN 230
D 626 KRDNLNLDSDHFDYDNGNLEIHTQLANDLDDLKQALNNANVKGIIVDQDQIEHID 695
QY 231 KITANNNGNDIRNLKFADELDVR--LYN-QELVERWNLAASDILRISMLKEDGGVYLDV 287
D 686 VSDANGKVIPTKRLANNNDPKNIIVTLKVVILKQNNQPNIDLISEEQLSGDNHISFK- 744
QY 288 DILPGIQDLPKSNKPD-----SITNTSWEMIKLEAMKVEYIPGYSKQNFMLDVEVQ 343
D 745 -----KPTITAKTKENDDYEISFSPNSLANKKIKLFTKDD-----NNTNTKTVEAS 791
QY 344 RSPESALSCKSKSEIFLP-----LDDIKV-----SPL----- 371
D 792 IGLDGKALFKTSDDAIFAPDHKTYLTKEADNKKVANIIDEISLDRIVNKKQKGNVWADN 851
QY 372 --EVKIAPANNSVI-----NQALISLSDSYCSOLIVNQIKNRY-----KI-- 409

```

```

Db 852 KHEFKIPDQKNKOLTAVYKDKNNNEIHVPIKTDKQKVIIVNPNNUFLDPNKIYDFDKIVD 911
QY 410 LNDNLNPSINEGDTFNTMKIFSD-----KLASISN--EDNMFMKIKITNVLKVG 457
D 912 LNEYPNKTILDRNSINKDVSAINDGVQAKVLKAPAVSNVTVAINFQVNLFDNLKLS 971
QY 458 ---FAPDVRSTINLSGPGVYTGAYQ-----DLMFKN 487
D 972 YNRQFALTIKAVNDLNDTKVIATYDPTNNYKLNFDTHLDANTKYKVVVDVLELNNK 1031
QY 488 STNHLLEPELRNFEFPKTKISQTEQETSLWSFNQARAKSQ-----FEYKKGVEG 541
D 1032 EKPIKLIKDDVLNFEFTSSATTIN---PPIWTKFDVVVTKTNNDDTTITFEIDDK---DN 1084
QY 542 ALGEDDNLDPQNTVLKDYVKKILSSMKTRKEYIHYIVQLQGDKISYEASCNLPSKD 601
D 1085 ILKNDQKIT-YAQLALMDDDDLDDTVDIMPLVNFST-----NKIASINGLDDKNS 1133
QY 602 PYSSILYQKNEGSETAYVYVADAIEK-----EIDKYRIPYQISKNRNIKLTFFIGHGK 655
D 1134 KYSI-----KNL-----YYLNDQNKVYLFKNDVTKYE-QHFITNPKINLSF--NK 1177
QY 656 SEFNTDTF---ANL-----DVDSLSEIEFETILNLAKADISPKY 690
D 1178 SAVEQDIPADHANLFDYKDYDQKLINEDVKIYQINIDNTKNELO---IGYKV-VANNK 1234
QY 691 IEINILGCMFYSISABEETPGKLLKIKDRVSELNPS-----ISQDSITV 737
D 1235 IKFNLVG-----LKEKTYTVIKKLEBALNKSASSIVNSEFDDLDPTNFTSNKNTLV 1287
QY 738 SANQY-----EVRINEEGKEIILDHSGKI--NKEESIIDKISSEY 777
D 1288 GLNSIDVWNDQTFIINAKINIGDDFQDNQOVKLIYVSDNKNKEIKSAVTLIKQNVYQF 1347
QY 778 ISFNPKNKIIVKSKY-----LHELSTLLOEI---RNA-----NSSDIDLEKVM 821
D 1348 EFSNLIKRLVTFKSIYVETNNQTLHKDLTLTHQFSINPNSNAVSLKNTNIEITKRLIV 1407
QY 822 TECEINVASNDROIVEGRIB--EAKNL--TSDSINYIKNEFKLIESISDSLYDLKHQNG 877
D 1408 -----NNDQSLISAKIEVDIDNVLNTADKFNIVYQ-----LENSDANNLKLATN 1454
QY 878 L-----DSDHISFEDI-----SKTENGFRIRFINKETGNSIETETKE 916
D 1455 AQOVVNNKKFLKFDVLSKLNQNVYIKESFNKPTNAY--FNFTNNKNTNIVSYDQON 1513
QY 917 IFSEVATHISKEISNIKDTIFDNV-----GKLVKVNLDAAHEVNTL----- 959
D 1514 -----KISLNNINPTSYKPKQTINKNDKSVNIDVDLQVDKQLLANQVLR 1557
QY 960 -----NSAFFIOSLIEYNTTKES--LSNLSVAMKVQVYAQLFS-TGLNTITDASKVV 1008
D 1558 LKQLKQNDKVTWTDPILFNNNAKISPKLSNL-----IHNRAVEGLVYFDDQNSVN 1610
QY 1009 ELVSTALDETDLLPTLSEGLPIATIIDGVSGLNAIKELS-----ETNDPLARQ 1058
D 1611 DMTNNQISFNSKIHKPKIEFEPSLTTI--NYDTNNAIKTVSAHNAQVHFUKTNDDEALN 1668
QY 1059 E-----IEAKIGIMAVNLTAASTAIVTSALGIASGFSILLVPLAGTS 1100
D 1669 DQIVBAVFAPTNLNDQKVVEAKLNNVNSFNEGELEFNLSGLREETYRLIKVTFKNKP 1728
QY 1101 AGIPSILVNN-----ELIQDKATKVIDYFKHISLABTEGAFTLDD----- 1141
D 1729 NKAYELLKNGVIFPKNGSQAYEFTQKFEKVIDVSVSSTNTTQOEITVKIDGIQRA 1788
QY 1142 ----KIIMPQDVLVSE-----IDFNNSI-----TLGKCEI 1169
D 1789 MNNKLELVYESNILGDPKIEKTTVDNNNKSVDHSPDKKEYNLVNLNLPKFGRRYSKKINI 1848
QY 1170 WBAEGSGHTLDDI--DHFF-----SSPSITYRKPWLSIYDVZNLKKEKIDFS 1216

```

Db 1849 KEVNGQDHEFVFNVNSFDVNLQSEITASSVEEINDRAP-----DKLNQTTIKINLK 1903
QY 1217 KDLVLPNAPNRVFGYEMGTGPFGRSLDNDGTGKLLDRI-RDHYEGQFYWRYFAPFIADALI 1275
Db 1904 DONDILK-----INDIATITYDNEQKQDAIVKTNAQOKY-----LEALI 1943
QY 1276 TKLKFYEDTNVRINLDGNTSFIVPVIITEQIRKNLSYFSGSGSYSLSLSPYNNVI- 1334
Db 1944 TNL-----VFNKWIIKKIEFKNLSQTF-----IKVGKNTNVI 1977
QY 1335 ----DLN--LVENDTWI-----DVD---NVVKI-----ITIESDEIQGELIENILSKLN 1376
Db 1978 YDESNLKLIINNDPQIIGPLSTDVNSTONVANKKHVISTLDFKYNPHISKNLKFKLK 2037
QY 1377 IE-----DNKIIINN--HTINFYGDINESNRFSLT-PSILEDINIIIEIDL 1420
Db 2038 FQNINGEVVYSPILTNSIIVNNKNVINFTDLNLSNQLYRLVDVYIDDNN-----DT 2092
QY 1421 VSKSYKILLSGNCKMLIE-----NSSDIQOKIDH--- 1449
Db 2093 INDKKVPKANNVTPIIDIAPKTISKSNNTWTSTSSQFVFNKSDGNEVLNDLEA 2152
QY 1450 -IGFNGEHQKYPYSDINETKNGFIDYSKKEGLTAEFNSIIRNIYMPDSNNLFY 1508
Db 2153 TISFKKGOTLLTPVKV--NIIKQNN--KYLKQGITTLEPENRYVLESILLAKEN----- 2203
QY 1509 SSKDLKDRIINKGV--KLLIGNYFKDDMKVSLFTIEDNTI--KLVGVLDENGVAQ 1564
Db 2204 KTKPLVVEILNKDDISFQOAGNYKVIQIKSQNPSTVDTKQIKLKLDDGIONAWNEKQL 2263
QY 1565 ILKEVNNAKSALNTSNNMFLESINIKNIF--YNNLDPN-----TEFILDN----- 1610
Db 2264 EITISAND-----NSTKTAIIKLEKVKLEYEFELTNLEKRYTPTKIELINDNNTKTPFN 2319
QY 1611 -----FIISGSNSIGQFEL--ICDKDKN-IQPYFINFKIETSITYLVGNRQNLIVE 1659
Db 2320 KSDSIQDKFIVLSNNQVGVGNIIEIQDRDVNHLNSAKIRFELNDLDNVL--SNDEQATIT 2377
QY 1660 PSYHLDDSGNISSTVINFSQKLYGI-----DRYVNVK 1692
Db 2378 YNNNOTTSAKV---ITDQNKYLEATFSNLVINKDTIINKI 2415

Search completed: April 1, 2004, 16:55:00
Job time : 84.875 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 16:39:30 ; Search time 25.625 Seconds
(without alignments)
3424.942 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1700
Perfect score: 8675
Sequence: 1 NMLVNKAQLQKMWVKFRIQ.....YLYGIDRYVNVKVIAPNLYT 1700

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/5A COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/5B COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/5A COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/5B COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6759.5	77.9	2366	1	US-08-480-604A-10
2	6759.5	77.9	2366	2	US-08-405-496A-10
3	6759.5	77.9	2366	3	US-08-915-136-10
4	6759.5	77.9	2366	4	US-08-957-310-10
5	6759.5	77.9	2366	4	US-10-011-366-10
6	6759.5	77.9	2366	4	US-09-084-517-10
7	4467	51.5	2710	1	US-08-480-604A-6
8	4467	51.5	2710	2	US-08-405-496A-6
9	4467	51.5	2710	3	US-08-915-136-6
10	4467	51.5	2710	4	US-08-957-310-6
11	4467	51.5	2710	4	US-10-011-366-6
12	4467	51.5	2710	4	US-09-084-517-6
13	512.5	5.9	3169	4	US-09-453-702B-257
14	347	4.0	10182	4	US-09-134-001C-3159
15	334	3.9	2807	4	US-09-543-681A-4980
16	315	3.6	2777	4	US-09-543-681A-6124
17	304	3.5	2184	4	US-09-417-485D-6
18	289	3.3	3433	4	US-09-091-501B-10
19	279	3.2	3696	4	US-09-134-001C-5080
20	265.5	3.1	2662	4	US-09-595-684B-31
21	265.5	3.1	3248	1	US-08-353-700-1
22	265.5	3.1	3248	5	PCT-US95-16216-1
23	263	3.0	2504	4	US-09-328-352-5821
24	259	3.0	2285	4	US-09-308-375-2
25	254.5	2.9	3135	1	US-08-323-170B-2
26	254.5	2.9	3135	4	US-08-954-441-2
27	253.5	2.9	2391	2	US-08-446-855A-2

Sequence 2, Appli
Sequence 1, Appli
Sequence 28, Appl
Sequence 28, Appl
Sequence 28, Appl
Sequence 28, Appl
Sequence 27, Appl
Sequence 1, Appli
Sequence 8, Appli
Sequence 1, Appli
Sequence 23, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 23, Appl
Sequence 50, Appl

ALIGNMENTS

RESULT 1
US-08-480-604A-10
; Sequence 10, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPD-01763	
TELECOMMUNICATION INFORMATION:	
TELEPHONE: (415) 705-8410	
TELEFAX: (415) 397-8338	
INFORMATION FOR SEQ ID NO: 10:	
SEQUENCE CHARACTERISTICS:	
LENGTH: 2366 amino acids	
TYPE: amino acid	
TOPOLOGY: linear	
MOLECULE TYPE: protein	
US-08-480-604A-10	
Query Match 77.9%; Score 6759.5; DB 1; Length 2366;	
Best Local Similarity 76.6%; Pred. No. 0;	
Matches 1302; Conservative 191; Mismatches 206; Indels 1; Gaps 1;	
QY	1 MNLVNAQQLQKVVYKFIQDEYVAIILNLEEYHNMSESSVVEKYLKLDINLTDNYL 60
DB	1 MSLNRKQLEKMANVRFQDEYVAIILDALEEYHNMSENIVVEKYLKLDINSLTDIYI 60
QY	61 NTKYKSGRNKALKPKPEYLTWEVLELKNLSLTPVEKMLHFIWIGQINDTAINYNQWD 120
DB	61 DTKYKSGRNKALKPKPEYLTWEVLELKNLSLTPVEKMLHFIWIGQINDTAINYNQWD 120
QY	121 VNSDYTKVYVDSNAFLINTLTKTVESATNTLTLESFRENLDPEFYNYKFKEMEIIY 180
DB	121 VNSDYNVNVFDSNAFLINTLTKTVESAINDTLLESFRENLDPRFYNYKFKRMEIIY 180
QY	181 DKQKHPIDYKSOIEENPEFIIDNIITKYLNEYSKOLEALNKYIEESLNKITANNNDI 240
DB	181 DKQKFINFYKQARENPELIDDIIVKTYLSNEYSKEIDELNTVIEESLNKITQNSGNDV 240
QY	241 RNLKFADEDLVRLNQELVERWNLAAASDILRISMLKEDGGVYLDVDILPGIQDLPKS 300
DB	241 RNFEFPKNGESFNLYQEQLVERWNLAAASDILRISALKEIGMYLDVDMPLPGIQDLPES 300
QY	301 INKPSITNTWEMIKLEAIKMYKEYIFGYTSKQFMDLDEEVQSFSALSKSKDKSEIF 360
DB	301 IEKPSVTVDPEWMTKLEAIKMYKEYIPEYTSHEFMDLDEEVQSFSFVSLASKDKSEIF 360
QY	361 LPDLIKYSPLEKTAFAANSVINQALJSLKDSYCSDLVINQIKRYKILNDNLNPSINE 420
DB	361 SSLGMEASPLEKTAFAKSGIINGLISVKDSYCSNLIVQIENRYKILNNSLNPAISE 420
QY	421 GTDFNTMTKIFDKLASINEDNMFMKITNYLKVGFAPDVRSTINLSGPGVVTGAYQD 480
DB	421 DNDFTNTTFDLSMAEANDANGFMELGKYLAVGFDPVKTTINLSGPEAYAAVQD 480
QY	481 LLMFKDNSTNIHLBEPENFPFKTKISQTEQBITSLWSPNOARAKSQPEEYKKGIFYE 540
DB	481 LLMFKEGSNTHLIEADLRNFEISKTINSQTEQEMASLWSPDDARAKAQPEEYKKNIFYE 540
QY	541 GAIAGEDNDLDFQNTVLDKDYYSKILSSMKTRNKEYIHYIYVQLQDKISYEASCNLFESK 600
DB	541 GSIAGEDNDLDFSQNIYVDVDEYLEKISSLARSSEYIHYIYVQLQDKISYEAAACNLFAK 600
QY	601 DPYSSITLYKNTBEGSTAYVYVADAEIKIYKRYPIYOISNKRNIKLTFFIGHGKSEFNT 660
DB	601 TPYDVSILFQKNIEDSIAIYYPGDEIGQIEIKYKIPSIISDRPKILTFIGHGKDEFT 660
QY	661 DTPANLDVDSLSSEIETILNLAKADISPKYIIEINLLGCMNFYSISASBETYPKILLKIK 720
DB	661 DIFAGFDVDSLSIEEADLAKEDISPKSIEINLLGCMNFYSINVBETYPKILLKVK 720
QY	721 DRYSELMPSISQDSITVSANQYEVRIINSEGRBLDHSKWNKINKEESIKDISSEYISF 780
DB	721 DKISELMPSISQDSITVSANQYEVRIINSEGRBLDHSGEWINKEESIKDISSEYISF 780
QY	781 NPENKTIYKSKYVHELSTLQIRNANSSDIDLEKKYMLTECEINVASNIDROIVEGR 840
DB	781 NPENKTIYKSKLPSTLLQIRNANSSDIDLEEKVMLTECEINVINSDITQIVEER 840
QY	841 IBEAKNLTSDSINVYKQEFKLIESIDALCDLKOQNELEDHSFISFEDISTDGGFIRF 900

RESULT 2
US-08-405-496A-10
; Sequence 10, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO

STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 424
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-405-496A-10

Query Match 77.9%; Score 6759.5; DB 2; Length 2366;
Best Local Similarity 76.6%; Pred. No. 0;
Matches 1302; Conservative 191; Mismatches 206; Indels 1; Gaps 1;

Qy 1 MSLVYKQLEKMANVRETOEDEYVAILDALEEHYHNSSENTVVEKYLKLDINSLDIYI 60
Db 1 MSLVYKQLEKMANVRETOEDEYVAILDALEEHYHNSSENTVVEKYLKLDINSLDIYI 60
Qy 61 NTKYKSGRNKALKKFKKYLTVLEVLELKNLSLTPVEKNLHFIWGGQINDTAINYINQWKD 120
Db 61 NTKYKSGRNKALKKFKKYLTVLEVLELKNLSLTPVEKNLHFIWGGQINDTAINYINQWKD 120
Qy 121 VNSDYTVKVFYDSNAFLINTLTKTIVESATNTLTSPRENLNDPEPDYKFKRMEIYY 180
Db 121 VNSDYTVKVFYDSNAFLINTLTKTIVESATNTLTSPRENLNDPEPDYKFKRMEIYY 180
Qy 181 DKQKHFDIYKQSEENPEFIIDNIKTYLSNEYSKOLEALNKYIEESLNKTIANNNGDI 240
Db 181 DKQKHFDIYKQSEENPEFIIDNIKTYLSNEYSKOLEALNKYIEESLNKTIANNNGDI 240
Qy 241 RNLEKFADEDLVRIYNGELVERWNLAAASDILRISMLKEGGVYLDVDDILPGIQLPLFKS 300
Db 241 RNLEKFADEDLVRIYNGELVERWNLAAASDILRISMLKEGGVYLDVDDILPGIQLPLFKS 300
Qy 301 INKPDSTINTSWEMIKLEAIKMYKEYIPGYTSKNPDMLEDEVORSEPSALSSKDKSEIF 360
Db 301 INKPDSTINTSWEMIKLEAIKMYKEYIPGYTSKNPDMLEDEVORSEPSALSSKDKSEIF 360
Qy 361 LPLDDIKVSPLEVKIAPANNVINQALISLKDSCSLVINQIKRKYKILNDNLNPSINE 420
Db 361 LPLDDIKVSPLEVKIAPANNVINQALISLKDSCSLVINQIKRKYKILNDNLNPSINE 420
Qy 421 GTDFNTMTKIFSDKLASINEDNNMFMKITYNLKVGFPDVRSTINLSGFGVYTGAYQD 480
Db 421 GTDFNTMTKIFSDKLASINEDNNMFMKITYNLKVGFPDVRSTINLSGFGVYTGAYQD 480

Db 421 DNDFTNTTNTFIDSIMAEANADNGRFNVELGKYLRVGFFDFDVKTTINLSGPEAYAAAYQD 480
Qy 481 LLMPKDSNTIHLLEPELRNFEPFKTKISQTEQEIITSLMSFNQARAKSFEEYKGYFE 540
Db 481 LLMPKDSNTIHLLEPELRNFEPFKTKISQTEQEIITSLMSFNQARAKSFEEYKGYFE 540
Qy 541 GALGEDNDLPAQNTVLDKQYVSKLLSMKTRNKEYIHVIVQLOQDKISYEASCNLPSK 600
Db 541 GALGEDNDLPAQNTVLDKQYVSKLLSMKTRNKEYIHVIVQLOQDKISYEASCNLPSK 600
Qy 601 DPYSILYOKNIEGSETAYYYYVADAEIKEDKRIYQVQISNKENIKLTIHGHKSSENT 660
Db 601 DPYSILYOKNIEGSETAYYYYVADAEIKEDKRIYQVQISNKENIKLTIHGHKSSENT 660
Qy 661 DTEANLDVDSLSSEIETILNLAADISPKYIEINLLGNMFSYSISABEYTPGKLLKIK 720
Db 661 DTEANLDVDSLSSEIETILNLAADISPKYIEINLLGNMFSYSISABEYTPGKLLKIK 720
Qy 721 DRVSELMPISIQDSITTVSANQYEVRIINEEGKREILDHSGKWINKESIIKDISKEYISF 780
Db 721 DRVSELMPISIQDSITTVSANQYEVRIINEEGKREILDHSGKWINKESIIKDISKEYISF 780
Qy 781 NPKNKIIIVKSKYLHELSTLLOETRNNANSDDIDLEKKVMLTECEINVASNIDQIVEGR 840
Db 781 NPKNKIIIVKSKYLHELSTLLOETRNNANSDDIDLEKKVMLTECEINVASNIDQIVEGR 840
Qy 841 IEEAKNLTSDSINYIKQEFKLIIESISDLYDLKHONGLDDSHFISFEFISDXTENGFRIRF 900
Db 841 IEEAKNLTSDSINYIKQEFKLIIESISDLYDLKHONGLDDSHFISFEFISDXTENGFRIRF 900
Qy 901 INKETGNSIFETEKEIFSEYATHISKEISNIKOTIFDNVNGKLVKKNLDAAEVNTLN 960
Db 901 INKETGNSIFETEKEIFSEYATHISKEISNIKOTIFDNVNGKLVKKNLDAAEVNTLN 960
Qy 961 SAFFIQSLIEYNTTKESLNSLVAMKVQVYAQLFSTGLNTITTDASKVVELVSTALDETID 1020
Db 961 SAFFIQSLIEYNTTKESLNSLVAMKVQVYAQLFSTGLNTITTDASKVVELVSTALDETID 1020
Qy 1021 LLPTLSEGLPIIATIIDGVSLGAAIKELSETNDPILRQEIETAKGIMAVNLTAASTAVT 1080
Db 1021 LLPTLSEGLPIIATIIDGVSLGAAIKELSETNDPILRQEIETAKGIMAVNLTAASTAVT 1080
Qy 1081 SALGHTASGFSILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETGAFTLLD 1140
Db 1081 SALGHTASGFSILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETGAFTLLD 1140
Qy 1141 DKIIIMPQDDLVLSEIDFNNSITLCKCEIMRAEGSGHLLTDDIDHFFSPSITIRKPWL 1200
Db 1141 DKIIIMPQDDLVLSEIDFNNSITLCKCEIMRAEGSGHLLTDDIDHFFSPSITIRKPWL 1200
Qy 1201 SIYDVLNLIKKEKIDFSDKLMVLPNAPRVFGYEMGWTGPFGRSLNDGFKLLDRIDHYEG 1260
Db 1201 SIYDVLNLIKKEKIDFSDKLMVLPNAPRVFGYEMGWTGPFGRSLNDGFKLLDRIDHYEG 1260
Qy 1261 QYTWYFAFIADALITKLPRIEDNTVINLDGNTRSPFIVPITTEQTKNLSYFSGG 1320
Db 1261 QYTWYFAFIADALITKLPRIEDNTVINLDGNTRSPFIVPITTEQTKNLSYFSGG 1320
Qy 1321 GSYLSLSPYNNIDNLAVENDTWVIDVNVVKNITTESDBIIOKGLTENILSKLINBDN 1380
Db 1321 GSYLSLSPYNNIDNLAVENDTWVIDVNVVKNITTESDBIIOKGLTENILSKLINBDN 1380
Qy 1381 KIILNNHNTINFYGDINESNRFTISLTFSILEDINIIIEIDLVSYSKYILLSGNCKLIENS 1440
Db 1381 KIILNNHNTINFYGDINESNRFTISLTFSILEDINIIIEIDLVSYSKYILLSGNCKLIENS 1440
Qy 1441 SDIOQKHIGHGNBHQKVIPIYSYIDNETKNGFIDYSKKGLFTABESNESIIRNIYMP 1500
Db 1441 SDIOQKHIGHGNBHQKVIPIYSYIDNETKNGFIDYSKKGLFTABESNESIIRNIYMP 1500
Qy 1501 DSNNTFIYSKDKLDRIINKGDKVLLIGNYFKDKMKVLSLFTIEDTITKLNGVYLDEN 1560
Db 1501 DSNNTFIYSKDKLDRIINKGDKVLLIGNYFKDKMKVLSLFTIEDTITKLNGVYLDEN 1560
Qy 1501 DSKPSFGYSSNNLKDVKVITKDNVNILTYGVLKODIKISLSLTLDQDEKTIKILNSVHDES 1560
Db 1501 DSKPSFGYSSNNLKDVKVITKDNVNILTYGVLKODIKISLSLTLDQDEKTIKILNSVHDES 1560

MOLECULE TYPE: protein
US-08-915-136-10
Query Match 77.9%; Score 6759.5; DB 3; Length 2366;
Best Local Similarity 76.6%; Pred. No. 0;
Matches 1302; Conservative 191; Mismatches 206; Indels 1; Gaps 1;
QY 1 MNLVKAQLOKVVYKPKIQDEYVAIINALBEEYHNMSESSVEKYKLDKIDNNLTNYL 60
Db 1 MSLVNRKQLEKMANVFRQDEYVAIILDALBEEYHNMSENIVVEKYKLDKIDNNLTNYL 60
QY 61 NTYKSGRNKALKPKFXYLMEVLELKNLSLTPVEKNLHFTWIGQINDTAINYNQWKD 120
Db 61 DTYKSGRNKALKPKFXYLMEVLELKNLSLTPVEKNLHFTWIGQINDTAINYNQWKD 120
QY 121 VNSDYTVKVFYDSNAFLINTLTKTIVESATNNTLESFRENLDNPEFVKNYKVRKEMEIIY 180
Db 121 VNSDYNVNVFYDSNAFLINTLTKTIVESAINDLESFRENLDNPEFVKNYKVRKEMEIIY 180
QY 181 DKQKHFIDYKSOIENPEFIIDNIKTVLSNEYSKOLEALNKYIEESINKITANGNDI 240
Db 181 DKQKHFIDYKSOIENPEFIIDNIKTVLSNEYSKOLEALNKYIEESINKITANGNDI 240
QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVDILPGIQDLPKS 300
Db 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVDILPGIQDLPKS 300
QY 301 INKPSITWSEMIKLEAIMKYIPIGYTSKNTFMDLDEEVQSFESALSCKSKSEIF 360
Db 301 INKPSITWSEMIKLEAIMKYIPIGYTSKNTFMDLDEEVQSFESALSCKSKSEIF 360
QY 361 LPDDIKVSPLEVKAFANNSVINQALSLKDSYCSDLVINOIKRYKILNLDNLPISNE 420
Db 361 LPDDIKVSPLEVKAFANNSVINQALSLKDSYCSDLVINOIKRYKILNLDNLPISNE 420
QY 421 GTDENTMTKIFSDKLASISNEDNMFMKITYLVKGFAPDVRSTINLSGGVYTCAYQD 480
Db 421 GTDENTMTKIFSDKLASISNEDNMFMKITYLVKGFAPDVRSTINLSGGVYTCAYQD 480
QY 481 LLMFKDSTNIHLLEPELRNFPFKTKISQLEQETSLWSNQARAKSQFPEYKKGYPF 540
Db 481 LLMFKDSTNIHLLEPELRNFPFKTKISQLEQETSLWSNQARAKSQFPEYKKGYPF 540
QY 541 GALGEDNDLDFRAQNTVLDKDYVSKILLSMKTNYKEYIHYIYVLOQDKTISYEASCNLF 600
Db 541 GALGEDNDLDFRAQNTVLDKDYVSKILLSMKTNYKEYIHYIYVLOQDKTISYEASCNLF 600
QY 601 DYSSTLYOKNIBEGSETAYVYVADAEIKEDKIRIPYOISNKRNIKLTFIGHKSEFNT 660
Db 601 DYSSTLYOKNIBEGSETAYVYVADAEIKEDKIRIPYOISNKRNIKLTFIGHKSEFNT 660
QY 661 DTFANLDVDSLSSEIETILNADKADISPKYIEINLLGCMFYSISAEETYPGKLLKTK 720
Db 661 DTFANLDVDSLSSEIETILNADKADISPKYIEINLLGCMFYSISAEETYPGKLLKTK 720
QY 721 DRYSELMPISQDSITVSANQYEVNRNEEGKREILDHSGKWNKESIIKDISSKEYISF 780
Db 721 DRYSELMPISQDSITVSANQYEVNRNEEGKREILDHSGKWNKESIIKDISSKEYISF 780
QY 781 NPENKIIVKSKYLHELSTLLOEIRNNANSSDIDLEKKWMLTECEINVASNIDRQVEGR 840
Db 781 NPENKIIVKSKYLHELSTLLOEIRNNANSSDIDLEKKWMLTECEINVASNIDRQVEGR 840
QY 841 IEEAKNLTSDSINVYKNEFKLIESISDLSYDLKHQGLDDSHFISFEDISKTENGFRIRF 900
Db 841 IEEAKNLTSDSINVYKNEFKLIESISDLSYDLKHQGLDDSHFISFEDISKTENGFRIRF 900
QY 901 INKETGNSIFETEKEIFSEYATHISKEISNIDKTFIDNVNNGKLVKKNLDAAEVNTLN 960
Db 901 INKETGNSIFETEKEIFSEYATHISKEISNIDKTFIDNVNNGKLVKKNLDAAEVNTLN 960
QY 961 SAPPFQSLIEYNTTKESLSNLSVAMKVQVYAOIPLSTGLNITITASKVVELVSTALDETID 1020
Db 961 SAPPFQSLIEYNTTKESLSNLSVAMKVQVYAOIPLSTGLNITITASKVVELVSTALDETID 1020

QY 1561 GVAQILKPMNAKALNTSLSMNFLESINIKNFYNNLDNPNTIEFILDTPFIISGNSIG 1620
Db 1561 GVAQILKPMNAKALNTSLSMNFLESINIKNFYNNLDNPNTIEFILDTPFIISGNSIG 1620
QY 1621 QFELICDNDKNIQPIFYKFKETSYLYVGNRQNLIVPEFSYHLLDSDGNSISSTVINFSQK 1680
Db 1621 QFELICDNDKNIQPIFYKFKETSYLYVGNRQNLIVPEFSYHLLDSDGNSISSTVINFSQK 1680
QY 1681 YLYGIDRYNVKVIAPNLYT 1700
Db 1681 YLYGIDRYNVKVIAPNLYT 1700
QY 1690 YLYGIDSCVKNVWISPIY 1699
Db 1690 YLYGIDSCVKNVWISPIY 1699
RESULT 3
US-08-915-136-10
Sequence 10, Application US/08915136
Patent No. 6290960
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESS: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,136
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/480,604
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear

961	AA	FF	IO	SL	IE	YN	SK	ES	SL	SN	SV	AM	QV	VQ	VA	QF	SL	GN	TI	TD	PA	KV	EL	SV	EL	TA	DE	TI	1020																			
1021	LL	PT	SE	GI	PI	IA	TI	ID	GS	LG	AA	IK	EL	SE	TD	PL	LR	QE	EA	IG	MA	VN	LT	AA	ST	AI	VT	1080																				
1021	LL	PT	SE	GI	PI	IA	TI	ID	GS	LG	AA	IK	EL	SE	TD	PL	LR	QE	EA	IG	MA	VN	LT	AA	ST	AI	VT	1080																				
1081	SA	LG	IA	GS	PI	LV	PI	AG	IS	AG	PS	LV	NN	EL	II	OD	KAT	VI	DY	PK	HS	LA	ET	SG	AT	LLD	1140																					
1081	SS	LG	IA	GS	PI	LV	PI	AG	IS	AG	PS	LV	NN	EL	II	OD	KAT	VI	DY	PK	HS	LA	ET	SG	AT	LLD	1140																					
1141	DK	II	MP	QD	DL	VI	SE	ID	FN	NN	SI	TL	GK	CE	IT	RA	SG	GG	SH	TL	TD	DI	HH	FF	SS	PI	TY	KP	WL	1200																		
1141	DK	II	MP	QD	DL	VI	SE	ID	FN	NN	SI	TL	GK	CE	IT	RA	SG	GG	SH	TL	TD	DI	HH	FF	SS	PI	TY	RE	PH	1200																		
1201	SI	YD	VL	NI	KK	EK	IF	DK	LM	VL	PA	NP	RV	PG	YEM	GT	PG	RS	LD	NG	TK	LL	DR	IR	DH	YEG	1260																					
1201	SI	YD	VL	EV	QK	EL	DL	SK	LM	VL	PA	NP	RV	PA	WT	GT	PG	RS	LE	NG	TK	LL	DR	IR	DH	YEG	1260																					
1261	QF	YR	YF	AF	IA	DAL	IT	KL	PR	YED	TN	VR	IN	LD	GN	TR	SF	IV	PI	TT	QE	IR	KN	LS	SF	YV	GS	1320																				
1261	EF	YR	YF	AF	IA	DAL	IT	KL	PR	YED	TN	VR	IN	LD	GN	TR	SF	IV	PI	TT	QE	IR	KN	LS	SF	YV	GS	1320																				
1321	GS	YS	LS	PS	PN	MM	ID	LN	VEND	TW	VD	VN	VN	KN	IT	ES	DE	IO	KG	EL	IN	IL	SK	LN	IEDN	1380																						
1321	GT	YAL	SL	QY	MM	GIN	EL	SE	DY	II	VD	VN	VD	VT	IES	DK	KG	DL	IE	GL	TS	IE	EN	1380																								
1381	KI	LN	NHT	IN	FY	GD	INES	NR	FI	SL	TS	ILED	INI	TE	IL	DA	VS	KY	KILL	SG	NC	KL	ENS	1440																								
1381	KI	LN	SH	IN	FS	GE	VGS	NG	FV	SL	TS	ILE	GIN	AI	TE	VD	LS	KY	KULL	IS	GE	KI	LM	NS	1440																							
1441	SD	I	Q	Q	K	I	D	H	I	G	P	GE	HQ	KY	I	P	Y	I	D	N	E	T	KY	NG	F	I	D	Y	S	K	KE	G	L	T	A	F	S	N	E	S	I	R	N	I	Y	M	P	1500
1441	NH	I	Q	K	I	D	I	G	P	NS	E	L	Q	K	N	I	P	S	F	V	D	S	E	G	K	E	N	G	F	I	N	G	1500															
1501	DS	NN	L	F	Y	S	K	OL	K	D	R	I	IN	K	G	V	D	K	L	L	I	GN	F	D	D	K	M	K	V	S	L	FT	IED	TN	T	K	L	NG	V	Y	D	EN	1560					
1501	DS	K	P	S	F	G	Y	NN	L	K	V	K	VI	T	Q	N	V	N	IL	T	G	Y	L	K	D	I	K	I	S	L	T	L	Q	B	E	K	T	I	K	L	NS	V	H	LD	EN	1560		
1561	GV	A	IL	K	FW	N	AK	S	AL	T	NS	L	N	KL	ES	I	N	I	K	N	I	F	YN	LD	PN	IE	F	L	D	T	N	F	I	L	SG	NS	I	G	1620									
1561	GV	A	E	IL	K	FW	N	R	-	K	NT	S	D	SL	S	F	LE	S	N	I	K	S	I	F	V	N	LO	S	N	I	K	F	I	D	AL	N	F	I										

RESULT 4

US-08-957-310-10
; Sequence 10, Application US/08957310
; Patent No. 6365158
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; APPLICANT: Kink, John A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
; TITLE OF INVENTION: DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medler & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; City: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

Db 541 GSLGDDNLDPSQNIWDPKEVLEKISSLARSSRGYIHYVQLQGDKISYEACNLFAK 600
 QY 601 DRYSSILYQKIEGSETAYYYVADAEIKEDKRYIPYOISKNENIKLTFIGHCKSEFNT 660
 Db 601 TPYDVLFPQKIEDSEIAYYVPGDGEIQEDKYPISIIIDRPKIKLTFIGHCKDEFT 660
 QY 661 DTFANLDVDSLSSETITINLAKADISPKYTEINLGCNMFYSISAEETVPGKLLKIK 720
 Db 661 DIFAGFDVDSUSTEAEADIAKEDISPKYTEINLGCNMFYSINVEETVPGKLLKVK 720
 QY 721 DRVSELMPSISQDSITVSQANQYEVRAINBEGKREILDHSGKWINKEEIIKDISKEYISF 780
 Db 721 DKISSELMPSISQDSIIIVSQANQYEVRAINBEGKREILDHSGKWINKEEIIKDISKEYISF 780
 QY 781 NPKENKIIVKSYLHELSTLOEBIRNNANSDDILEKKWMLTECEINVASNIDQIVEGR 840
 Db 781 NPKENKITVKSXNLPESLTLQEBIRNNANSDDILEEKVWMLTECEINVISNIDQIVEER 840
 QY 841 IEEAKNLTSDSINYIKNPFKTESISDYLDKHONGLDDSHFISPEDISKTENGFRIRF 900
 Db 841 IEEAKNLTSDSINYIKNPFKTESISDALCDKQONELEDSHFISPEDISKTENGFRIRF 900
 QY 901 INKGTGNSIFITEKEIFESEYATHISKEISNIKOTIFDNVAGKLVKYNLDAAEHVNTLN 960
 Db 901 INKGTGESIFVETEXTIFSEYANHITSEISKIGTIFDTVAGKLVKYNLDTTHEVNTLN 960
 QY 961 SAFFIQLSEYNTKTESINLGVAMKVQYVQLFSTGLTNTIDAKVVELVSTALDETID 1020
 Db 961 AAFIFIQLSEYNTKTESINLGVAMKVQYVQLFSTGLTNTIDAKVVELVSTALDETID 1020
 QY 1021 LPTLSEGLPIIATIDGVSIGAAIKELSETNDPLLRQIEIAKIGIMAVNLTAASTAIVT 1080
 Db 1021 LPTLSEGLPIIATIDGVSIGAAIKELSETNDPLLRQIEIAKIGIMAVNLTAASTAIVT 1080
 QY 1081 SALGIASGSIILLVPLAGISAGIPSLVNNELLQDKATKVIDYFKHISLAETGAFTLLD 1140
 Db 1081 SGLGIASGSIILLVPLAGISAGIPSLVNNELLVDRKATKVIDYFKHISLAETGAFTLLD 1140
 QY 1141 DKIIIMPQDVLVSEIDFNNSITLKGCEIWRAGGSGHTLTDIDHFFSSPSITVYKPEWL 1200
 Db 1141 DKIIIMPQDVLVSEIDFNNSITLKGCEIWRAGGSGHTLTDIDHFFSSPSITVYKPEWL 1200
 QY 1201 SYDVUNIKKEKIDPSKLMWIPNAPNVRVGVEMGWTGFRSLNDGDKLDRDRHYEG 1260
 Db 1201 SYDVUNIKKEKIDPSKLMWIPNAPNVRVGVEMGWTGFRSLNDGDKLDRDRHYEG 1260
 QY 1261 QFYWRVFATADALITLKPREDTNRINLGNSTRSFIVPITTEYREKLSYSFYSGG 1320
 Db 1261 QFYWRVFATADALITLKPREDTNRINLGNSTRSFIVPITTEYREKLSYSFYSGG 1320
 QY 1321 GSYSLSPYNNIDNLVENDTWIDVNVYKNITIESDEIQGELIENILSKLINBDN 1380
 Db 1321 GYVALSLOYNMGINIELSESDWIIDVNVVDRVTIESDKIKGDLLEGILSTLSIEN 1380
 QY 1381 KIILNHTINFYGDINESNRFLSFLPSILEDINIIEIDLVSKYKILLSGNCKMLIENS 1440
 Db 1381 KIILNHSHEINFSGEVNGSVFLSFLPSILEGINAIEVLLSKYKILLSGELKILMNS 1440
 QY 1441 SDIOQKIDHIGFNGEHOQIPYSYIDNETKNGFDIDYSKKEGLFTABPSNESIIRNIYP 1500
 Db 1441 NHIOQKIDYIGFNSQLQKIPYSYIDNETKNGFDIDYSKKEGLFTABPSNESIIRNIYP 1500
 QY 1501 DSNULFIYSKDLDIRIKNGDVKLLIGNYFKDDMKVSLFTIEDTNTIKLNGVYLDEN 1560
 Db 1501 DSKPSFGYNNKDKVKTOKVNTLTGYIKDDIKISLSTLQDEKTIKLSVHLDEN 1560
 QY 1561 GVAQILKFMNNAKSALNTSNLMFLIESINIKNIFYNNDPNIETFLDTNFIISGNSIG 1620
 Db 1561 GVAELIKFMNR-KGNTNTSDSLMSFLESWKIKSIFVNFQSNIKPLDANFIISGTSIG 1619
 QY 1621 QFELICOKKNOIOPYFINFKIETSVTLYVGNRQNLIVPSVHLDGSGNISSTVINFSQK 1680
 Db 1620 QFEFICDENDNOIOPYIKFNTLTETNYTLVYVGNRQNMIVPNYDLDGSGDISSTVINFSQK 1679

QY 1681 YLYGIDRVYKVIAPNLVT 1700
 Db 1680 YLYGIDSCVKNVISPNIYT 1699
 RESULT 5
 US-10-011-366-10
 ; Sequence 10, Application US/10011366
 ; Patent No. 6573003
 ; GENERAL INFORMATION:
 ; APPLICANT: Williams, James A.
 ; Kink, John A.
 ; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
 ; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
 ; DISEASE
 ; NUMBER OF SEQUENCES: 22
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSES: Medlen & Carroll
 ; STREET: 220 Montgomery Street, Suite 2200
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: United States of America
 ; ZIP: 94104
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/011,366
 ; FILING DATE: 16-No. 6573003-2001
 ; CLASSIFICATION: <unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/957,310
 ; FILING DATE: 23-OCT-1997
 ; APPLICATION NUMBER: US 08/329,154
 ; FILING DATE: 24-OCT-1994
 ; APPLICATION NUMBER: US 08/161,907
 ; FILING DATE: 02-DEC-1993
 ; APPLICATION NUMBER: US 07/985,321
 ; FILING DATE: 04-DEC-1992
 ; APPLICATION NUMBER: US 07/429,791
 ; FILING DATE: 31-OCT-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Ingolia, Diane E.
 ; REGISTRATION NUMBER: 40,027
 ; REFERENCE/DOCKET NUMBER: OPHD-01121
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 705-8410
 ; TELEFAX: (415) 397-8338
 ; INFORMATION FOR SEQ ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2366 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
 US-10-011-366-10
 Query Match 77.9%; Score 6759.5; DB 4; Length 2366;
 Best Local Similarity 76.6%; Pred. No. 0;
 Matches 1302; Conservative 191; Mismatches 206; Indels 1; Gaps 1;
 QY 1 MNLVNAQLQKQVYKFRIOEDVEVALLNALEEVHNSSESVVEKYLKLDINLTDNYL 60
 Db 1 MSLVNRKQLSEKQANVRFTQEDYVAILDALEEVHNSSESVVEKYLKLDINLTDIYI 60
 QY 61 NTYKSGRNKALKKFKFELTMEVLELKNLSLTVEKNLHFVITCGQINDTAIYINQWKD 120
 Db 61 DTYKSGRNKALKKFKFELTMEVLELKNLSLTVEKNLHFVITCGQINDTAIYINQWKD 120
 QY 121 VNSDYTVKVFYDSNAPLINTLTKTIVESATNNTLESFRENLDPEFDYKFRMEIYY 180

Db 121 VNSDYNVFPYDANFLNLTKTIVVESAINDTLESFRENLDNPRDYNNFKRMEIY 180
Qy 181 DKQHFIDYKQIEENPEFIDNIIKTYLSNEYSKDLALNKYIBESLNKITAANGNDI 240
Db 181 DKQKNFINYKQAQREENPELLIIDIVKTYLSNEYSKIEDELNTYIBESLNKITAQSGNDV 240
Qy 241 RNLEKFADELDVLYNQELVERNLAAADILIRISMLKEDGGYLDVDILPGIQDLPKS 300
Db 241 RNPEEFKNGESFNLYEQELVERNLAAADILISALKSIGMYLDVMDLPGIQDLPES 300
Qy 301 INKPDSITNTSWMIKLEAMKYKEYIPGYTSKNFMDLDEEVORSFESALSSKSKSEIF 360
Db 301 IEPSSVTVDFWMTKLEAMKYKEYIPEYTSHEFMDLDEEVOSSESVLASKSKSEIF 360
Qy 361 LPDDIKVSPLEVKIAPANNVINQALISLKDSCYCDLVINQIKNRYKIINDNLPSINE 420
Db 361 SSGJDEASPLEVKIAPNSKGIINOGLISVKDSCYCNLVKIQIENRYKIINSLNPAISE 420
Qy 421 GTFNTMTKIFSKLASISNEDNMFMKITNYLKVGFAPDVRSTINLSGPGYVYQAYOD 480
Db 421 DNFNTTNTFIDSIMAEANADNGRFMWELGKYLVRGVFPDVKTITNLSGPEYAYAAQD 480
Qy 481 LLMFKONSTNHLLEPELRNPEPKTKISQITTEBITSLMSFNQARAKSQPEEYKGYFE 540
Db 481 LLMFKEGSMNHLIEADLRNFEIKTNISQSTECEMASLWSDFAKAKAQFEYKKNYFE 540
Qy 541 GALGEDDNLDPQONTVLDKDVSKILLSMKTRNKEYIHYIYVLOQDKI-SYASCNLPSK 600
Db 541 GSGJEDDNLDPQONVVVDKYLEKISSLARSGYIHYIYVLOQDKI-SYEAACNLPAK 600
Qy 601 DPXSSILYKQIEGSETAYYYVADAEIKEDKIRIPYQISNKRNIKLTFFIGHKSEFNT 660
Db 601 TPYDSVLFOKNIEDSEIAYYVNGDGEIQEIDKVKIPIISDRPKIKLTFFIGHKDEPNT 660
Qy 661 DTFANLDVDSLSSEIETILMAKADISPKYIEINLLGCNMFYSISABETYPGKLLLKX 720
Db 661 DIFAGFDVDSLSSEIEAIDLAKEDISPKSIEINLLGCNMFYSINVEETYPGKLLLKX 720
Qy 721 DRYSELMPISQDSITVSANOYEVNNEBGRKILDHSGKWINKESIIKDISKEYISF 780
Db 721 DKISELMPISQDSITVSANOYEVNNEBGRKILDHSGKWINKESIIKDISKEYISF 780
Qy 781 NPENKIIVSKYLHELSTLQETRNANSDDIDLEKKVMLTECEINVASNIDRQVEGR 840
Db 781 NPENKIIVSKYKLPSTLQETRNANSDDIDLEKKVMLTECEINVASNIDRQVEGR 840
Qy 841 IEEAKNLTSDSINYKNEFKLIESDSLYDLKHQGLDSDHSPISFEDISKTENGPRIRF 900
Db 841 IEEAKNLTSDSINYKDEFKLIESISDALCDLQKQNELEDSDHSPISFEDISKTENGPRIRF 900
Qy 901 INKETGNSIETETEKEIFSEYATHISKEISNIDKTIPTDNVNGKLVKXNLDAAHEVNTLN 960
Db 901 INKETGESIFVETEKTIIFSEYANHITBEISKIKGTIPTDNVNGKLVKXNLDAAHEVNTLN 960
Qy 961 SAPIQSLIEVNTKESLSNLVAMQVQVYAOVAFSLGTLNITDASKVVELVSTALDETID 1020
Db 961 AAPFIQSLIEVNTKESLSNLVAMQVQVYAOVAFSLGTLNITDASKVVELVSTALDETID 1020
Qy 1021 LLPTLSEGLPIATIIDGVSLGAIKELSETNPLLRQIEAKIGIMAVNMLTAATAIVT 1080
Db 1021 LLPTLSEGLPIATIIDGVSLGAIKELSETNPLLRQIEAKIGIMAVNMLTAATAIVT 1080
Qy 1081 SALGIASGFSILLVPLAGISAGIPSLVNNELLIQDKATKVIDYFKHISLAETEGFTLLD 1140
Db 1081 SSLGIASGFSILLVPLAGISAGIPSLVNNELLVLRDKATKVVDYFKHISLAETEGFTLLD 1140
Qy 1141 DKTIIMPDDIVLSIEDFNNSITLGKCEIWRAGSGSHTITDDIDHFFSPSITYKPKWL 1200
Db 1141 DKTIIMPDDIVLSIEDFNNSITLGKCEIWRAGSGSHTITDDIDHFFSPSITYKPKWL 1200
Qy 1201 SIYDVLNKKKIDFSDKDLVLPNPNRVFAMETGTPGLRSLDNDGTCLLDRIIRDHYEG 1260

Db 1201 SIYDVLVQKEELDLSKDLVLPNPNRVFAMETGTPGLRSLDNDGTCLLDRIIRDHYEG 1260
Qy 1261 QFYRYFAFADALITKLPREDYTNIRINLDGNTRSFIVPVTTEIRKNLSYSFYGSG 1320
Db 1261 EFWRYFAFADALITKLPREDYTNIRINLDGNTRSFIVPVTTEIRKNLSYSFYGSG 1320
Qy 1321 GSYLSLSPYNNMTDLNLVENDTWIDVNVVKNITTESDEIOKGLIENILSKLNIEDN 1380
Db 1321 GTYALSLSQYNGMINIELSDVDWIIDVNVVRDVTTIIESKIKKGDLEGLILSLISEEN 1380
Qy 1381 KIILNHTINPYGFINESNRFISLTSILEDINIIEIDLVSYSKYILLGNCMKLIENS 1440
Db 1381 KIILNSHEINFSGEVNGSNGFVSLTFSILEGINAIEVDLLSKYKLLISGELKILMNS 1440
Qy 1441 SDIOQKIDHIGFNGEHOKYIPYSVIDNETKNGFIDYSKKEGLTAFESNESIIRNIMP 1500
Db 1441 NHIQOKIDYIGFNGSELQKNIPYSFVDSGKENGFGINGSTKEGLFVSELPDVLISKVMYD 1500
Qy 1501 DSNLPIYSSKDLKDIRIINKGDVKLLIKNYFKDDMKVLSFTIEDNTIKLNGVILDEN 1560
Db 1501 DSKPSFGYSNLDKDVKITKONVNILUTGYLKKDDIKISLTLQDEKTIKINSVHLOES 1560
Qy 1561 GVAQILKFMNNAKALNTSLSMNFLESINIKNIFNNLDPNIEFILDTPFIISGNSIG 1620
Db 1561 GVAELKFMNR-KGNTNTSDLSMSFLESMAIKSIFVNFLOSNIKIFILCANFIISGTSIG 1619
Qy 1621 QFEILCDKXNIQYFNFINKIKETSYTYLVGNRQNLIVBPSYHLDDSGNISSTVINFSQK 1680
Db 1620 QFEICDENDNIQYFNFINKIKETSYTYLVGNRQNLIVBPSYHLDDSGNISSTVINFSQK 1679
Qy 1681 YLYGIDRVNKKVIAPIHLYT 1700
Db 1680 YLYGIDSCWKNKVISPIY 1699

RESULT 6
US-09-084-517-10
; Sequence 10, Application US/09084517
; Patent No. 6613329
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/084,517
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: OPND-01610
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-084-517-10

Query Match 77.9%; Score 6759.5; DB 4; Length 2366;
Best Local Similarity 76.6%; Pred. No. 0;
Matches 1302; Conservative 191; Mismatches 206; Indels 1; Gaps 1;
QY 1 MNLVNAQLOKQWVVKRIQDEVAIINALAEYHNHNSSESVVEKYLKLDINLNDNYL 60
DB 1 MSLVNRKOLEKXANVRFTQDEVAIIDLAEYHNHNSSESVVEKYLKLDINSLDIYI 60
QY 61 NTYKSGRNKALKKFKYLYLMEVLELKNNSLTPVEKNLHFIWGGQNDTAINYNQWKD 120
DB 61 DTYKSGRNKALKKFKYLYLMEVLELKNNSLTPVEKNLHFIWGGQNDTAINYNQWKD 120
QY 121 VNSDYTVKVFVDSNAFLINTLTKTIVESATNTNTLESFRENLDPEPDYKFKYKMEIY 180
DB 121 VNSDYTVKVFVDSNAFLINTLTKTIVESATNTNTLESFRENLDPEPDYKFKYKMEIY 180
QY 181 DKQHFIDYKSOETEENPEFIIDNIITKYLSEYNSKOLEALNKYIEESLNKITANNENDI 240
DB 181 DKQKFNINYKAQREENPELLIDIVKYLSEYNSKIDELNTYIEESLNKITQNSGNDV 240
QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGVYLDVDIIPGQDLPFS 300
DB 241 RNFEFKNGESFNLYEQELVERWNLAAASDILRISALKEIGGMVLDVDMPLGGIOPDLFES 300
QY 301 INKPSINTNWEMLKLEAMKYKEYIPGYTSRNFMDLDEBQVRSFESALSSKSDKSEIF 360
DB 301 IEXPSVTVDFWEMTKLEAMKYKEYIPEYTSRNFMDLDEBQVRSFESALSSKSDKSEIF 360
QY 361 LPLDIDKYSPLEKIAFANNVINQALISLSDSCVSLVINQIKVRYKILNDNLNPSINE 420
DB 361 SSLGDMFASPLEVKIAFNSKGIINQGLISVSDSCVSLVINQIKVRYKILNLSNPAISE 420
QY 421 GTDRNTWKIFSDKLASISNEDNMFMKITNLYKVGFPADPVRSTINLSGPGVYTGAYQD 480
DB 421 DNDFTNTTNTFIDSIMAZANADNGRFMMELGKYLVRGFFPDVKTINLSGPEAYAAAYQD 480
QY 481 LLMFKXDSNTNHLLEFELNPEFPKTKISQTEGITSLSNFSNQAARAKSOFEEYKGYFE 540
DB 481 LLMFKEGNSNHLIEADLRNFEISKTNISQTEGEMASLWFSFDDARAKAOFEEYKRYFE 540
QY 541 GAGGEDNDLPAQNTVLDKDVSVKILSSMTKRYEYHYIYVQLQDKISYEASCNLPFSK 600
DB 541 GSLGEDNDLDFSQNTVWDKEYLLEKISSLASRSESGYIHYIYVQLQDKISYEAAACNLF 600
QY 601 DPYSILYQKNLEGSETAYVYVADABEIKEDKYRIPQISKNRNIKLTFIGHCKSEFNT 660
DB 601 TPYDSVLFOKNLEDEIAYYNGPGEIQEIDKYKIPSIISDRPKIKLTFIGHCKDEFNT 660
QY 661 DTFANLVDLSLSSETITLAKADISPKYTEINLLGCNMFYSYSABEYTPGKLLLIK 720
DB 661 DIFAGFDVLSLSEIAALDAKEDI SPKSEINLLGCNMFYSYSINVEYTPGKLLIKVX 720
QY 721 DRVSELMPISQDSITVGSANQYEVINEEGREILDHSGKWINKEESIIDKISSKEYISF 780
DB 721 DKISELMPISQDSITVGSANQYEVINEEGREILDHSGKWINKEESIIDKISSKEYISF 780

RESULT 7
US-08-480-604A-6
; Sequence 6, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.

QY 781 NPKENKIIVKSKYLHELSTLLOEIRNNANSSDIDLEKKVWLTECEINVASNIDQIVEGR 840
DB 781 NPKENKIIVKSKNLPSTLLOEIRNNSSDIDLEKKVWLTECEINVISNIDQIVEER 840
QY 841 IEEAKNLTSDSINYIKNEFKLIESDSLYDLKHQGLDSDSHFISFEDISTENGFRIRF 900
DB 841 IEEAKNLTSDSINYIKNEFKLIESDALCDLQKQNELESDSHFISFEDISTENGFRIRF 900
QY 901 INKETNSIFITEKEKIFSEYATHISKEISNIKOTIPDNVNGKLVKKNLDAAEHVNTLN 960
DB 901 INKETGESIFVETEKTIIFSEYANHITBESIKIGTIFDTVNGKLVKKNLDTTHEVNTLN 960
QY 961 SAFFIQSLIEYNTTKESLSNLVAMKQVYAQLFSTGLNTITDASKVVELYSTALDETID 1020
DB 961 AAFPIQSLIEYNSKESLSNLVAMKQVYAQLFSTGLNTITDASKVVELYSTALDETID 1020
QY 1021 LPLTSEGLPIIATIDGVSLSGAAAKELSETNDPLLRQIEIAKIGIMAVNLTAATAIVT 1080
DB 1021 LPLTSEGLPIIATIDGVSLSGAAAKELSETSDPLLRQIEIAKIGIMAVNLTTATTAIT 1080
QY 1081 SALGIASGFSILLVPLAGISAGIPSLVNNELILODKATKVDYFKHISLAETEGAFILLD 1140
DB 1081 SSLGIASGFSILLVPLAGISAGIPSLVNNELVLRDKATKVDYFKHISLVEGVTILLD 1140
QY 1141 DKINPQDVLVSEIDFNNSITLCKEIRWAEAGSGHTLTDDIDHFFSSPSITYRKEPL 1200
DB 1141 DKINPQDVLVSEIDFNNSITLCKEIRWAEAGSGHTLTDDIDHFFSAPSITYREPHL 1200
QY 1201 SIYDVLNLIKKEIDPSKOLMVLNPNRNVFGYEWGTPGPRSLDNDGTJLDRDRHYEG 1260
DB 1201 SIYDVLVQKEBELDLSKOLMVLNPNRNVFAMEGTPGLSLDNDGTJLDRDRHYEG 1260
QY 1261 QFYMYFAFIADALITKLPRYEDTNVRLNDGNTRSFIVPVTTEQIRKNSLYSFYSG 1320
DB 1261 EFYMYFAFIADALITKLPRYEDTNVRLNDGNTRSFIVPVTTEQIRKNSLYSFYSG 1320
QY 1321 GYSISLSPYNNIDNLVENDTWIDVNVVKNITTESDIOKGLIENILSKLNIEDN 1380
DB 1321 GYALSLSQYNNGINIELSESDVMIIDVNVVRDVTIESDKIKKGLIEGILSTLSEBEN 1380
QY 1381 KIILANNTINFGVDINESRFTSLTFSILEDINIIEIDLVSYSKILLSGNCKMLIENS 1440
DB 1381 KIILNSHEINFGVNGSGFVSLTFSILEGINALIEVDLLSKYKLLISGELKILMUNS 1440
QY 1441 SDIQOKIDHIGNGSHQYIPYYSYIDNETKNGFYDYSKKEGLFTAEPFNSIIRNIYMP 1500
DB 1441 NHIQOKIDYIGNSBLQXNIPYSFVDSKENGFGNGSKGLFVSELPDVVLISKVYMD 1500
QY 1501 DSNNLFIYSSKDLDIRIINKGDVKLLIGNYFKDDMKVSLSTFTEDTNTIKNGVYLDEN 1560
DB 1501 DSKPSFGYYSNNLKVYITKDNVNILGYLKKDDIKISLTLQDEXTIKLNSVHLDDES 1560
QY 1561 GVAQILKFNNAKSAALNTSLSLNFLESINIKVFNLDNPIEILDTNFIISGNSIG 1620
DB 1561 GVAELIKFMNR-KGNTNTSDLSMFSLESMTKISIFWNFLQSNIKFILDANFIISGTTIG 1619
QY 1621 QFELICDKXN:QPYFIPFKIKETSITLYVGNRQNLIVEPSYHLDSDGNSISTVINFSQK 1680
DB 1620 QFELICDENDN:QPIFKENTLETNTLYVGNRQNLIVEPYVLDSDGDISSTVINFSQK 1679
QY 1681 YLYGIDRYVNVKVIIAPNLYT 1700
DB 1680 YLYGIDSCVNVKVISPIY 1699

APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-604A-6
Query Match 51.58; Score 4467; DB 1; Length 2710;
Best Local Similarity 50.78; Pred. No. 2.2e-262;
Matches 871; Conservative 366; Mismatches 440; Indels 42; Gaps 15;
QY 1 MNLVNAQLQKVVYKFRIOEDYVAINALAEYHNHNSSESVVEKYLKLDINNLTNYL 60
DB 1 MSLISKELIKLAY-SIRPRENEVKTILNLDYENKLTNNENKYLQKLKINESIDVFM 59
QY 61 NTKYKSGRNKALKKFKYLVMEVLELKNLSLTPVEKNLHFITWGGQINDTAINYNQWKD 120
DB 60 NKYTKSRNALSNNLKDKLKEVILIKNSNTSPVEKNLHFVWIGGEVSDIALEYIKQWAD 119
QY 121 VNSDYTKVFDVSNAPLINTLTKKTIVESATNTLSEFRENLDPEEDYNKFKYKRWELIY 190
DB 120 INAEYNIKLWDSFAVLNLTAKAIVSSSTTEALQLLEESIQNPQFNNKFKYKRWELIY 179
QY 181 DKQKHFDYKQSEENPEFTIDNIITKYSNEYSKDLKALNKYIESLNKTIANNNDI 240

180 DRQKRFINYYKQINKPTVPTIDDIKSHLVSEYNRDETVEISYRTNSLRKINSNHDGI 239
QY 241 RNLEKPADEDLVRLYNQELVERWNLAAASDILRLSMLEKEDGGVYLDVLLPGIOPDLFKS 300
DB 240 RANSLFTEQELLNIYQOELLNRGNLAAASDIVLLALKNPFGVYLDVMDLPGHSHDLFKT 299
QY 301 INKPDSTINTSWEMIKLEAMKYKEYPGYTSKNFDMLEDEEVQSPESALSSKSDKSEIF 360
DB 300 ISRPSISGLDRWEMIKLEAMKYKYNNTSNTFDFKLDQOLKDNFKLIIESKSEIF 359
QY 361 LPLDDIKVSPLEVKIAPANNVINQALISLKDSCYSDLVINQKRYKILNDNINPSINE 420
DB 360 SKLENLVSDLEIKIAPALGALVINQALISQSGVLTNLVIEQVKRYOFLNQHLPATIES 419
QY 421 GTDNTNMLKPSDKLASISNEDNMFMKITNVLKVGFPADPVRSTINLSPGVYTGAYOD 480
DB 420 DNNFTDTKIFHDSLFNSATAENSMTLKIAPILOVGFMEARSTLSLQSGFAYASAYD 479
QY 481 LLMEKDNSTNHLLEPELRNPEPKTKISOLTEQITSLWSFNQARAKSOFEEYKKGFFE 540
DB 480 FINQENTIEKTLKASDLIEFKFPENNLSQLTEQIENSLWSFQASAKYQPEKIVRYTG 539
QY 541 GALGEDDNLDPQANTVLDKDY-VSKILSS--MKTRNKYIHYIVQLQGDKISYEASCNL 597
DB 540 GSLSENGVDENKNTALDKYLLNNKIPSNVVEAGSKNYVHYIQLQGGDISYEACNL 599
QY 598 FSKDPYSIIYQKNIQEGSEYAYYYVAD--AEIKEDKYIPQVQISNKRNIKUTFIHGK 655
DB 600 FSKNPKSIIIQNNV--NESAKSYFLSDGSESILELNKYRIPERLKNKKEKVKVTFIHGK 657
QY 656 SEFNTDTFANLDVDSLSESEITILNADKADISPKYIEINLLGCMNPSYSISAEETYGKL 715
DB 658 DEFNTSEFARLSVDSLSEISPLDTIKDISPKNVEVNLGCMNPSYDFNVEETYGKL 717
QY 716 LLKIKDRVSELMPSISODSITVSANOYEVRINEEGREILDHSGKWNKEESIKDIDSSK 775
DB 718 LLSINDKITSLPDVNNKNSITIGANQVEVRINSEGRKELLAHSGKWNKEEAINMSDLSK 777
QY 776 EYISFNPKENKIIVKSKYLHELSTLLOEIRNANSSDIDLEKKVMLTECEINVASNIDRQ 835
DB 778 EYIFDSDIDNKLKAKSKNIPGLASISEDIKTLLDASVSPDTKFIANLKNLNISSIGDY 837
QY 836 IVEGRIBEAKNLTSDSINYKNPEKLESISDSIYDLKHONGLDDSHFIPEDISKTENG 895
DB 838 IYVEKLEPVKNIHNSIDDLIDEFNLENVSDLEYELKLNLDKYLISPEDISKNSST 897
QY 896 FRIRPINKETGNSIFIEKEIFSEYATHISKEISNTKTFIDNVNGKLKVKNLDAHE 955
DB 898 YSVRFINKSGESVYVETEKEIFSKYSEHTKEISTIKNSIITDVGNNLNDNIQDHTSQ 957
QY 956 VNTLNSAFFIOSLIETNTTKESLNSLVAMKVQVYAOQFSTGLNTITDASKVVELVSTAL 1015
DB 958 VNTLNAAFFIOSLIDYSSNKVDLNDLSTSVKQVYAOQFSTGLNTIYDSIQLVNLSNAV 1017
QY 1016 DETIDLLPTLSEGLPIIATIIDVSGLGAAIKELSETNDPLLRQRIKIGIMAVNLTAAS 1075
DB 1018 NDTINVLFTITEGPIVSTILDGINLGAAIKELDEHDLPLKKEAKVGLAINMSLSI 1077
QY 1076 TAVTSALGASGSPILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETEGA 1135
DB 1078 AATVASIVGIGAEVTFILLPITAGISAGIPSLVNNELILHDKATSVMVNFHLSSEKYPG 1137
QY 1136 FTLLDDKILMPQDDLVLSEIDFNNSITLKGCEIWRABGGSGSHLTDDIDHFFSPSPITY 1195
DB 1138 LKTEDDKILVPIDDLIVSEIDFNNSIKLGTGCTNIAIEGGSGHVTGIDHFFSPSPISS 1197
QY 1196 RKPMLSIYDVNLINKEKIDFSDKMLVLPNAPNVPFVGVEMGTGPFRLSDNDGTLLDRIR 1255
DB 1198 HPSLSIYSAGIETENLDFSKIMLPNAPSPVFWETGAVPGLRSLNDGTLLDRIR 1257
QY 1256 DHYEQFVYRPAFIADALITKLKRYEDTNRINLNDGNTSRSFIVPVTITQIRKNLSYS 1315
DB 1258 DLYPGKFVYRFVAF--DYAITTLKPVYEDTNIKIKDKDTRNFIMPTITTTNEIRNKLVS 1316

Qy	361	LPDDDIKVS	PLEVKIAF	ANNSVINQ	ALISLKD	SYCD	SLDVL	INQIK	KNRYKIL	NDNL	APSINE	420																																														
Db	360	SKLENV	NSDLE	KIAPALG	SVINQAL	ISKQ	SYLTNL	VLIEQ	VKNRYQ	FLNQH	NPATES	419																																														
Qy	421	GTDPNT	MTKIP	SKDLAS	ISNED	NMF	KITNTY	LVKGF	APDP	VRST	NLSPG	VYTGAYD	480																																													
Db	420	DNNFTD	TKTFH	SDLSF	NSA	TAENS	MFLT	KIAP	VLQGF	MPPEAR	STISL	SGP	AYAYD	479																																												
Qy	481	LLMPK	ONSTNI	HILLEP	LRN	FER	PKTKI	SOLT	CEITS	LSW	SFQ	AKAK	SQF	EYK	KGYE	540																																										
Db	480	FINLQ	ENTI	BEKT	LKAS	LI	EFK	PENN	LSOLT	CEIN	LSW	SFQ	AKAK	QF	EYK	VADYTG	539																																									
Qy	541	GALCED	NLDL	FAQNT	VLDD	KDY	--VSK	ILSS--	-MKTR	NKEV	IHYI	VLQ	GD	KISYE	ASCNL	597																																										
Db	540	GSLS	EDNG	VDFN	KNTAL	DKYLL	NNK	ITPS	NNVE	EAGSK	NNV	HHII	QLQ	DDISYE	ATCNL	599																																										
Qy	598	FSKDP	YSSILY	QKN	IEG	SETAY	YYYV	AD--	AET	KEID	KRIP	QOI	NNK	NIK	UTFI	GHOK	655																																									
Db	600	FSKN	PKNS	II	IQRM	--	NEG	AK	YFL	SD	DES	ILE	NK	VRP	ERL	NK	KEK	VKVT	FI	GHOK	657																																					
Qy	656	SEFNTD	TFAN	LVD	VS	LS	SE	ETIL	NL	AK	AD	IS	PKY	TE	IN	L	GC	N	M	F	SY	S	AB	ET	Y	PG	KL	715																														
Db	658	DEFNTS	EFAL	S	VD	S	UNE	IS	FF	LT	TK	LD	IS	PN	VE	N	L	GC	N	M	F	SY	D	N	VE	ET	Y	PG	KL	717																												
Qy	716	LLKIK	D	R	V	S	E	L	M	P	S	I	S	O	D	S	T	V	S	A	N	O	V	E	R	I	N	E	G	K	R	E	I	L	D	H	S	K	K	I	N	E	E	S	I	I	D	I	S	S	K	775						
Db	718	LLS	T	M	D	K	I	T	S	L	P	D	V	N	K	S	I	T	I	G	A	N	O	V	E	R	I	N	E	G	K	E	L	L	A	H	S	K	K	I	N	E	E	A	I	M	S	D	L	S	S	K	777					
Qy	776	EYIS	F	N	P	K	E	N	K	I	I	V	K	S	Y	L	H	E	L	T	L	O	E	I	R	N	N	A	N	S	D	I	L	E	K	K	M	U	T	E	C	E	I	N	V	A	S	N	I	D	R	Q	835					
Db	778	EYI	P	F	D	S	I	D	N	K	L	K	A	K	S	N	I	P	L	A	S	E	D	I	T	L	L	D	A	S	V	S	P	D	T	F	I	L	N	L	K	I	N	E	S	I	G	Y	837									
Qy	836	I	V	E	G	R	I	E	B	A	K	N	L	S	D	S	I	N	I	K	N	F	K	L	E	S	I	S	D	S	I	L	D	L	K	H	O	N	G	L	D	D	S	H	F	I	S	F	E	D	I	S	K	T	E	N	G	895
Db	838	I	Y	E	K	L	E	P	V	K	N	I	I	H	N	S	I	D	D	I	D	E	F	N	L	E	N	S	D	E	L	Y	E	L	K	N	N	D	E	K	I	L	S	I	F	E	D	I	S	K	N	K	N	S	T	897		
Qy	896	F	R	I	R	I	N	K	E	T	C	N	S	I	F	E	T	E	K	E	I	S	E	V	A	T	H	I	S	K	E	S	I	N	I	K	D	T	I	F	N	V	N	G	K	L	V	K	N	L	D	A	A	E	895			
Db	898	Y	S	V	F	I	N	K	S	E	V	Y	T	E	K	E	I	S	K	Y	S	E	H	I	T	K	S	I	I	T	K	S	I	I	T	D	V	N	G	N	L	L	D	I	Q	L	D	H	T	S	Q	957						
Qy	956	V	N	T	L	N	S	A	F	F	T	Q	S	L	I	E	N	T	K	E	S	I	N	S	V	A	M	K	V	Q	Y	A	O	L	F	T	G																					

```

1436 LIENSDDIOQIDHIGFNGEHQKYLIPYSYID-NETKYNGFIPIYSKEGILFTAEFSNESII 1494
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
1437 LISNLNTEIKENTGLD---SKNHAYNTDESNNKYFGAI-----SKTSOKSI 1483
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||
1495 RNITMPSDNL-----PIYSSXDL--KDRIINKGDVKLLIGNYPKD---DMKVSLIS 1541
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||
1484 H--YKDSKNILEFYNDSTLEFNKSDIAEDINVFMKODINIITGYVVDNNTDKSIDPS 1541
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||
1542 FTIEDNTIUKLVYLDENGVAQILKPMONAKSALNTSNLSLNFLBSINIKNIIFYNNLPD 1601
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||
1542 ISLVSKNQWVGVLYNLESYISSYLDPVKNSDGHNTSNFMLFLDNIISFWKLFGFE--- 1598
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||
1602 NIEFIELDTNEPISGSNSIGOFELICDKDKNIQBYFPNFKICTSYTLTYGNRLNLIPEPS 1661
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||
1599 NINPFVIDKYFTLVGNKLVNGVEFICDNNKNIDIYFGCEWKTSSKSIFSGNGENVVVEPI 1658
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||
1662 YHLDDSGNISSTVINFSQKLYXGIDRYVNKKVIAPNLYT 1700
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||
1659 YN-PDTGEDISTGLDFSVPPLYGIDRYINKVILIAPDLYT 1696
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||

```

RESULT 10

US-08-957-310-6
; Sequence 6, Application US/08957310
; Patent No. 6365158
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; APPLICANT: Kirk, John A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING
; TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR TI
; TITLE OF INVENTION: DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medien & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,310
; FILING DATE: 23-OCT-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 24-OCT-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-957-310-6

Query Match	51.5%;	Score	4467;	DB	4;	Length	2710;
Best Local Similarity	50.7%;	Pred. No.	2.2e-262;				
Matches	871;	Conservative	366;	Mismatches	440;	Indels	42;
Gaps	15;						
QY	1	MNLVNAQLOQKVVYKFRIQEIDEYVAILNALBEYHNMSSESVVEKYKLKOINNLTDNYL	60				
DB	1	MSLISKEELJKLAY-SIRPRENEYKTLTNLDEYNKLTNNNENKYVLOQKLNESIDVFM	59				
QY	61	NTYKSGRKNALKKFKFYITMVEVLEUKNNLSLTPVEKNLHFHFIWIGQINDTAINYNQWD	120				
DB	60	NKYKTSRRNALSNLKKDILKEVILIKNSNTSPVEKNLHFVWIGGEVSDIALEYIKQWAD	119				
QY	121	VNSDYTVKVPYDSNAELINTLKTIVESATNTNTLESFRENLDPPDYNKFKVKMEIYY	180				
DB	120	INAEYNIKULWSEAFVNTLKKAVESSTTALQVLEBEIQNPOFDNNKFKYKKRMEFIY	179				
QY	181	DQKHFIDYVKSQIENPFIIDNIITKVLNSNEYSKDLKALNKYIBESLNKTIANNNDI	240				
DB	180	DQKRXINYYKSOINKPTVPTIDDIKSHLVSEYNEDETVLSSYRTNSLRKINSNHGDI	239				
QY	241	RNLKFADEDLVELYNQCELVERWNLAASDIIRISMLKEDGGVYLDVDTLPGIQDPLFKS	300				
DB	240	RANSLFTEQELLNIYSOELLNRGNLAASDIVERLLALKNFVGYLDVDMPLGHSDLFKT	299				
QY	301	INKPDSITNTSEMILKLEALIMKYKEYIPGYTSKNPDMLEEVQVSPESALSXSCKSEIF	360				
DB	300	ISRPSSIGLDRWEMIKLEAIMKYKVINNYTSSNFDKLQDLKDNFKLIIESKSEKSEIF	359				
QY	361	LPJDDIKVSPLEVKIAPANNSVINOALISIKDSYCSDLVINQIKNRYKILNDNLPSINE	420				
DB	360	SKLENLVNDSLEIKIAPALGSVINQALISKQGSYLTNLVIEQVKNRYQFINQHLNPAIES	419				
QY	421	GTDFTNTMTKIFSDKLASISNEDNMFMKITNTYLVKGFAPVRSNTINLSGPGYVTGAYOD	480				
DB	420	DNNFTDTTKFHDLSNLSAENSMPITKIAPIVQGVPEARSTISLSPGAYASAYD	479				
QY	481	LMFMKONSTNIHLEPELRFEPFKYKISOLTSQBITSLWSFNQARAKSQPEYKKGYPE	540				
DB	480	FINQENTIEBTKLASDLIEFKFPENNLSQLTEQENSLWSFPQASAKYQFEKYVADYTG	539				
QY	541	GALGEDNDLPAQNTVLDKDY-VSKTLSS--MKTNRKEVIHVIVOLQGDKISYEASCNL	597				
DB	540	GSLSBNGVDNFKNATLDKNYLLNNKIPSNNVSEAGSNVHVIIQLQGGDISYEATCNL	599				
QY	598	PSKDPYSSILYQKNIEGSETAYVYVAD--AEIKEIDKVIPIYQIGNKNIKLFTFICHGK	655				
DB	600	PSKNPKNSIIITQRNM--NESAKYFYFSDGDESILELNKVIPELKNKKEKYVTFICHGK	657				
QY	656	SEFNTDTFANLVDLSLSEETIINLAKADISPKYIEINLLGCMFYSISABETYPGKL	715				
DB	658	DEFNTSEFARLSDVLSLNEISSFLDTTKLDJSPKVEVNLGCMFSDNFVEETYPGKL	717				
QY	716	LLKIKDRVBELMPSISODSTIVSANQVEYVRINSEGKKEIILDHGSKVINKEESIIKDISK	775				
DB	718	LLSIMDKITSTLPVNKNKSITIGANQVEYVRINSEGKKEILLAHSGKVINKEEAIMSLSK	777				
QY	776	EYISPNPKENKIIVKSKYLHELSTLLQEIERNANSSDIDLEKKVMLTECEINVASNIDRQ	835				
DB	778	EYIFPDSIDNKLKAKSKXNIPGLASISDDIKTLILLDASVSDPTKPLNNLKNIESISGDY	837				
QY	836	IVEGRIBEAKNLTSDSINYIKNPFKLESISDLYLKHONGLDDSHFISFEDISKTENG	895				
DB	838	IYIEKLEPVKNI IHNSIDDDIIDEFNLENVSDLEYELKKNLNDKYLISFEDISKNST	897				
QY	896	FRIRFINKETGNSIFITEKEIPSEYATHISKEISNIKOTIFDNVNGKLVKKNVNLDAHE	955				
DB	898	YSVREFINKSGESVYVETEKEIPSKYSEHITKEISTIKNSIITDVNGNLNDNLQDHTSQ	957				
QY	956	VNTUNSAFFTOSLIEYNTTESLNSLVAMKVQVYAOLFSTGLNTIITDASKVELVSTAL	1015				
DB	958	VNTUNAAFFIOLSIDYSGNKOVDNLSTSVKQVYAOLFSTGLNTIYDSIQLVNLISNAV	1017				

RESULT 11
US-10-011-366-6
; Sequence 6, Application US/10011366
; Patent No. 6573003
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; OF TOXIN A AND TOXIN B FOR THE TREATMENT
; DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOCS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/011.366
 FILING DATE: 16-Nov-2001
 CLASSIFICATION: <Unknown>
 PRIORITY DATA:
 APPLICATION NUMBER: US/08/957.310
 FILING DATE: 23-OCT-1997
 APPLICATION NUMBER: US/08/329.154
 FILING DATE: 24-OCT-1994
 APPLICATION NUMBER: US/08/161.907
 FILING DATE: 02-DEC-1993
 APPLICATION NUMBER: US/07/985.321
 FILING DATE: 04-DEC-1992
 APPLICATION NUMBER: US/07/429.791
 FILING DATE: 31-OCT-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Ingolia, Diane E.
 REGISTRATION NUMBER: 40,027
 REFERENCE/DOCKET NUMBER: OPHD-01121
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 705-8410
 TELEFAX: (415) 397-8338
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2710 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-10-011-366-6

Query Match 51.5%; Score 4467; DB 4; Length 2710;
 Best Local Similarity 50.7%; Pred. No. 2.2e-262;
 Matches 871; Conservative 366; Mismatches 440; Indels 42; Gaps 15;

QY 1 MNLVNAQLOKQVYKPIQEDVEYAILNALAEYHNMSESSVVEKYLKLDINLNTDYL 60
 DB 1 MSLISKEELIKLAY-SIRPRENEYKITLNTDENVKLTNNENKYLQLKINESIDVFM 59

QY 61 NTKKSGNKALKKPKYVLTVEVLKKNLSLTPVEKNLHFWIGQINDTAINTINQWKD 120
 DB 60 NKYKTSRNRALSNLKKILKEVILKNSLTPVEKNLHFWIGVSDIALEVIKQWAD 119

QY 121 VNSDYTVKVFYDSNAFLNTLKKTTIVESATNNTLSPRENLDPEFVKNYKRMELIY 180
 DB 120 INAEYNKILWYDEAFVNTLKAIVESSTTEALQLLEEEIQNPOFQNMKFKYKRMFIY 179

QY 181 DKQHFIDYKQOISENPEFIDNIKYLNEYSKDLKALNKYIEESLKNITANGNDI 240
 DB 180 DRQKRFINYKQINKPTVPTIDDIKSHLVSEYNRDETIVLESYRTNSLRKINSNHGIDI 239

QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILRLISMLKEDGVYLDVLDLPGIQDLPKS 300
 DB 240 RANSUFTQOELLNYSQELNLGNLAASDILVRLALKNFGVYLDVDMPLGIIHSDLPKT 299

QY 301 INKPSITNTSMEIKLEAIMKYKEYIPGYTSKNFMDLDEEYQRFESALSKSDKSEIF 360
 DB 300 ISRPSIGLDRMEMIKLEAIMKYKYINNTYSENFDKLDQDLKDNFKLIESKSEKSEIF 359

QY 361 LPDDIKVSPLEKIAFANNSVINQALSLKDSYCSGLDVINOIKRKYILNDNLNPSINE 420
 DB 360 SKLENLWSDLEIKIAFALGVSINQALISKQSGLYTLNVLIEQVKRKYQFLNQHLPALIES 419

QY 421 GTDFNTMTKIFSKLASINEDNMFMKITYLKVGFAPDVRSTINLSGGCVYTGAYQD 480
 DB 420 DNNFTDTTKIFHDSLFSNATANSPLTKIAPYLOVGFMPPEARSTISLSGGAYASAYD 479

QY 481 LAMFKDNSTNHLLEPELNFPPKTKISQLEQETISLWSFNQARAKSQPEEYKGYFE 540
 DB 480 FINLQNTYEKTLKASDLIEFKFEPENNLSQLTEQELNSLWSDQASAKYQEKYVRYDTG 539

QY 541 GALGEDNDLFAQNTVLDKDY-VSKKILSS--MKTRNKEYIHYIVQLQGDKITSYPASCNL 597
 DB 540 GSLSLEDNGVDNKTALDKNYLLNNKIPSNVVEAGSKNYVHYIIQLQGDIDISYEATCNL 599

QY 598 FSKDPYSSILYQKNIEGSETAYVYVAD--AEIKEIDKYRIPYQISNKRNIKLFIHGK 655
 DB 600 FSKPKNSIIIQRM--NESAKYFLSDDGESILELKYRIPERLKNKKEKVKVTFIHGK 657

QY 656 SEFNTDTFANLDVDSLSSEIETILNLAKADISPKYIEINLLGCNMFYSISAETYPCKL 715
 DB 658 DEFNTSEFARLSVDSLSNEISSFLDTIKLIDSPKNVEVNLGCNMFYDFNVEETYPCKL 717

QY 716 LKIKDRVSELMPSISQDSITVSANQYVRINEGKEILDHSKWTINKBESIIKDISSK 775
 DB 718 LLSIMDKITSTLPDVNKNISITIGANQYVRINSEGRKELLASHKWTINKBESIMDSLSK 777

QY 776 EYISFNPKENKIIKSKYLHELSTLQEIARNANSDDIDLEKKVMTLCEINVASNIDRQ 835
 DB 778 EYIFFSDINLKAQKNIPGLASISEDIKILLDASVPTKFIANLKLNISSIGDY 837

QY 836 IVEGRIEERAKNLTSDSINYIKNEFKLIESISDSYDLKHQGLDSDSHFISFEDISKTEG 895
 DB 838 IYVEKLEPVKNIHNSIDDLIDFNLLENVSDLEYELKLANLDEKYLISFEDISKNST 897

QY 896 FRIRINKETGNSIFIEFTEKEIFSEYATHISKEISNIKDTTFDNGVKLVKKNLDAAE 955
 DB 898 YSVRFINKSGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVGNLLDNIQDHTSQ 957

QY 956 VNTINSAPFIQSLIEYNTTKESLSNLSVMKVQVYQALFSTGLNTITDASKVZLVSTAL 1015
 DB 958 VNTLNAAPFIQSLIDYSSNKNVNDLSTSVKQVYQALFSTGLNTIYDSIQLVNLSNAV 1017

QY 1016 DETIDLPTLSEGLPIATIIDGVSIGAAIKELSETNDPLLRQETEAIGIMAVNLTAAS 1075
 DB 1018 NDTINVLPTITIGFIVSTILDGINLGAIAKELLEDHPLLKKELEAKVGVLAINNSLSI 1077

QY 1076 TAVTSALGASGFSILLVPLAGISAGIPSLVNNELIQQDKATKVIDYFKHISLAETGA 1135
 DB 1078 AATVASIVGIGAEVIFLLPIAGISAGIPSLVNNELIHDKATSVVNVFNHLSSEKYP 1137

QY 1136 FTLLDDKIIMPODDLVLSEIDFNNSITLGKCEIWRAGGSGHTITDDIDHFFSPSITY 1195
 DB 1138 LKTEDDKILVPIDDLVISIDFNNSIKLGTGNILAMEGGSGHTVTGNIDHFFSPSIS 1197

QY 1196 RPEWLSIYDVNLTKKEKIDFSKDLVLPNAPNRVGYEMGTPGFRSLDNDGTCKLDRIR 1255
 DB 1198 HPSLSIYSAIGIETENLDFSKKIMLNPAPGRVFWETGAVGLRSLDNDGTCLDRIR 1257

QY 1256 DRYEGOFYRPAFIADALITKLKPYEDTVNRIINDGNTRSFIVPVITTEQIRKNLSYS 1315
 DB 1258 DLYPGFYRPAFF-DYAITTLKPYEDTNKIKLKDTRNFIMPTITTEIRKNLSYS 1316

QY 1316 FYGSGSYSLSPYNNNIDNLVENDTWIDVNVNKNITTESDEIQGELIENILSKL 1375
 DB 1317 FDGAGGYSLLSSYPISNTNLSKDDLWIFNIDNEVBREISIENTGKKGKLIKDVLSKI 1376

QY 1376 NIDNKIILNNHTINFYGDINESNRISLTSLEIDNIIIEIDLVSYSKYLISGNCMK 1435
 DB 1377 DINKLIIGNQITDFSGIDNKNRYIFLTCLDDKISLIIIEINLVAKSYLLSGDKNY 1436

QY 1436 LIENSDDIOQKIDHIFNGEHOXYPIYSYID-NETKINGFIDYSKKEGLTAFFESNESII 1494
 DB 1437 LISNLNTIEKINTGLD---SKNIAYNTDSNNKYFGAI-----SKTSQKSI 1483

QY 1495 RNTYMPDSNNL-----FIYSKDL--KDIRINKGDVKLLIGNYFKD---DMKVSJS 1541
 DB 1484 H--YKDSKNLIFYNDSTLEFNSKDFIAEDINVMKDDINTITGKYVYDNNNTDKSIDFS 1541

QY 1542 FTIEDNTIKLNGVYLDENGVAQILKFMNNAKALNTSNLWNLFSINIKNIFFYNLDP 1601
 DB 1542 ISLVSKNQVKNGLYNLESVYSYLDVFKNSDGHNTSNFMNLFNLDNIFWKLFGFE--- 1598

QY 1602 NIEFILDNTFIISGNSIQOFELICDKDKNIQYEFNEFKETSYTYLVGNRONLIVEPS 1661
 DB 1599 NINFVIDKYFTLVGKTNLGYVEFIDNNKNIDIFYGEWKTSSKSTIFSGNGRNVVEPI 1658

QY 1662 YHDDSGNISTVINFQSKLYGDRYVKNVLIAPNLYT 1700
 Db 1659 YN-PDGEDISTSLDYSYPLGYIDRYINKVLIAPDLYT 1696

RESULT 12
 ; Sequence 6, Application US/09084517
 ; Patent No. 661329
 ; GENERAL INFORMATION:
 ; APPLICANT: KINK, JOHN A.
 ; APPLICANT: WILLIAMS, JAMES A.
 ; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
 ; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
 ; NUMBER OF SEQUENCES: 30
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
 ; STREET: 220 MONTGOMERY STREET, SUITE 2200
 ; CITY: SAN FRANCISCO
 ; STATE: CALIFORNIA
 ; COUNTRY: UNITED STATES OF AMERICA
 ; ZIP: 94104
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/084,517
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/
 ; FILING DATE: 16-MAR-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/329,154
 ; FILING DATE: 25-OCT-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/161,907
 ; FILING DATE: 02-DEC-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/985,321
 ; FILING DATE: 04-DEC-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/429,791
 ; FILING DATE: 31-OCT-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: CARROLL, PETER G.
 ; REGISTRATION NUMBER: 32,837
 ; REFERENCE/DOCKET NUMBER: OPD-01610
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 705-8410
 ; TELEFAX: (415) 397-8338
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2710 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-084-517-6

Query Match 51.5%; Score 4467; DB 4; Length 2710;
 Best Local Similarity 50.7%; Pred. No. 2.2e-262;
 Matches 871; Conservative 366; Mismatches 440; Indels 42; Gaps 15;

QY 1 MNLVNAQLQKQWVYKFRIGEDYVAINALAEYHNNSSSVYKYLKDKDINNLTNYL 60
 Db 1 MSLISKEELIKLAY-SIRPRENEYKTLTNDYVKNLTNNENKYLQLKKNESIDVFM 59

QY 61 NTYKSGRNKALKKFKYLYLMEVLELKNLSLTPVEKVLHFWITGGQINDTAINYNQWQD 120
 Db 60 NKYKTSRRNALSNNKDKILKEVILIKNSNTSPVEKVLHFWITGGQINDTAINYNQWQD 119

QY 121 VNSDYTVKVFYDSNAFLINTLTKTIVESATNNTLESFRENLDPEFDYKFKRMEIY 180
 Db 120 INAEYNIKWYDSEAFVNTLKAIVESSITTEALQLLEEIEIQNPQDMFKFYKCRMEFY 179
 QY 181 DKQKFTDYKYSQIEENPEFIIDNIKTYLSNEYSKOLEALNKYIEESLANKITANNNDI 240
 Db 180 DRQKRFINYKYSQINKPTVPTIDDIKSHLVSEYNREDTIVLESYRTNLSLRKINSHGIDI 239
 QY 241 RNLEKPADEDLVRLNQLVERWNLAASDILRISMLKEDGGVYLDVILPGLOPDLFKS 300
 Db 240 RANSLFTEQELLNTYSQELNLRGNAASDIVLLALNFGGVYLDVDMPLGHSOLFKT 299
 QY 301 INKPDSTINTSWEMIKLEAIMKYEIPGYTSKNFMDLDEEVORSFESALSSEKSEIF 360
 Db 300 ISRPSSIGLDRWEMIKLEAIMKYKYNNTYSENFKLDQQLKDNFKLIESKSEKSEIF 359
 QY 361 LPLDDIKVPLEVIAFANNVINQALISKDSYCSOLVINQKRYKILNNDLNPSINE 420
 Db 360 SKLENLVSDLEIKIAFALGVSINQALISKGSYLTNLVIEQVKNRYQFLNQLNPAIES 419
 QY 421 GTDENFTWKIFSDKLASINEDNMFMKITYLVKGFADPVRSTINLSGGVYTGAYQD 480
 Db 420 DNNFTDTTKIFHDSLFNSATAENSMTLKAPLYLQVGFMPKARSTISLSGPGAYASAYD 479
 QY 481 LLMEKDNSTNHLLEPELRNFERPKTKISOLTEQITSLWSFNQARAKSOFEEYKKGFFE 540
 Db 480 FINQENTIEKTLKASDLIEFPENNLSQTEQINSLSWSPQASAKYQFEKTVRYTG 539
 QY 541 GALGEDDNLDPQNTVLKDY-VSKILSS--MKTRNKEYIHYIVQLQGGKISYEASCNL 597
 Db 540 GSLSEDNGVDPNKNTALDKVLLNKNIPSNVBEAGSKNYVHYIIQLQGGDISYEATCNL 599
 QY 598 PSKDPYSILYQKNIEGSETAYYYVAD--AEIKEDKIRIPYQISKNRIKLTIFGHGK 655
 Db 600 PSKNPKNSIIIQRNW--NESAKSYFLSDGSESILELNKRYIPERLKNKKEKVKYVFIHGK 657
 QY 656 SEFNTDTFANLDVDSLSEIETILNLAADISPKYIENLLGCMFYSYSIAEITYPKL 715
 Db 658 DEFNTSEFARLSVDSLSEISFSDTLTKDISPKNVEVNLGCMFYSYDFNVEITYPKL 717
 QY 716 LLKIKDRVSELMPSISODSITVSANQYEVIRINEGKEILDHSGKWKINKEESIKDISK 775
 Db 718 LLSIMDKITSLPDVKNKNSITIGANQYEVIRINEGKELLAHSGKWKINKEEAINSDISSK 777
 QY 776 EYISFPNPREKIIIVKSKYLHELSTLOEIRNANSSDILEKKVMTLCEINVASNIDRQ 835
 Db 778 EYIFPDSIDNKLKAKSKNIPGLASISEDITKLLLDASVSPDTKFIILNKLNISSIGDY 837
 QY 836 IVEGRIBEAKNLTSDSINYKNBFKLIESDSYDLKHONGLDSDHIFSPEDISKTENG 895
 Db 838 IYVEKLEPVKNIIEHNSIDDLIDEPNLENVSDLEYELKLNLDKYLISPEDISKNST 897
 QY 896 FRIRFINKETGNSIFETEKEIFSEYATHISKEISNIKDTIFDNVNGKLVKKNLDAHE 955
 Db 898 YSVRFINKNGESVYVETEKEIFSKYSEHTKEISTKNSIITDVNGLLDNIOQDHTSQ 957
 QY 956 VNTLSAPFIOSLIETYNNTKESLSNLSVAMQVYVYQALFSTGLNTITDASKVVELVSTAL 1015
 Db 958 VNTLNAFFIOSLIDYSSNKDVNDLSTSVKQVYVYQALFSTGLNTIYDSIQVNLISNAV 1017
 QY 1016 DETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSETNDPLLRQRIEAKIGIMAVNLTAAS 1075
 Db 1018 NDTINVLPTITEGPIVSTILDGINLGAAIKELDEHDPLLKKELEAKVGLAINMLSLI 1077
 QY 1076 TAITVSAIGIASGFSILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISIAETEGA 1135
 Db 1078 AATVASIVGIAEVTIFLLPIAGISAGIPSLVNNELILHDKATSVVNYFNHLSKSKYGP 1137
 QY 1136 FTLDLDDKIIPQDDLVLTSEIDFNKNSITLKGCEIWRAGSGSHTLTDDIDHFFSSPSITY 1195
 Db 1138 LKTEDDKILVPIDDLVISEIDFNKNSIKIGTCNILAMEGSGHVTGIDHFFSSPSISS 1197
 QY 1196 RKPMLSIVDVLNKKKIDFSDKOLMVLNPNAPRVFGYEMGWTGPRSLDNDGTLLDRIR 1255

Db 1198 HIPSLISYSAIGITENLDFSKKIMLPNAPSRVFWHETGAVQLRSLNDGRLDSIR 1257
Qy 1256 DHEGQFYRYFAFIADALITKLPRYEDTNVRINLGNTRSFIVPVITTEQIRKNLSYS 1315
Db 1258 DLYPGKFYRYFAFF-DVAITLAPVYEDTNIKLQKDRNFIMPITITNEIRKNLSYS 1316
Qy 1316 FYGGGYSLSLSPYNNMIDNLVENDTWIDVNVKNITISDEIQKGLIENTLSKL 1375
Db 1317 FDGAGGTYSLLSSYPISNTINLSKDDLMIFNIDNEVREISIENTIKKGLIKDVLSKI 1376
Qy 1376 NIEONKILNHTNFYGDINESNRFSLTFSILEDNIIEIDLVSQSKILLSGNCKM 1435
Db 1377 DINKNKLIIGQIDFSGDINDKORYIFLTCEDOKLSLIEINLVAKSLLLSGKNY 1436
Qy 1436 LIENSDDQOQIDHIGFNGERHQKIPYSYID-NETKINGFIDYSKEGELTAFBSNESII 1494
Db 1437 LISNLSNTEIKNLGLD--SKNIAYNTDESNNKYFGAI-----SKTSQKSII 1483
Qy 1495 RNIMYDNNL-----FYSSKOL--KDIRINKGDKVLLIGNYFKD---DMKVSLS 1541
Db 1494 H--YKDSKNILEFYNDSTLEFNSKDFIAEDINVFMKDDINTITGKYVVDNNTDKSIDFS 1541
Qy 1542 FTIEDNTIKLVYLDENGVAQILKPNNAKSALNTSNLMFLSINIKNIFYNNLDP 1501
Db 1542 ISLVSKNQVNGLYLNEVSYSYLDVFKNSDGHHTSNFNNFLDNLISFWKLFGE--- 1598
Qy 1602 NIEFTLDNFTISGNSIGQELICDKDKNTQPFINFKIKETSITLYVGNQNLIVEPS 1561
Db 1599 NINVIDKYFLVGNKTLGYVEFCDDNKNNDIVFGWKTSKSTTFPSGNRVVVEPI 1558
Qy 1662 YHLDGSGNISSTVNFSGYLYGIDRYVKNVVIAPNLYT 1700
Db 1659 YN-PDTGEDISTSLDFSPELYGIDRYINKVLIAPDLYT 1696

RESULT 13

US-09-453-702B-257
; Sequence 257, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; Applicant: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000

TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 257:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3169 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 257:
US-09-453-702B-257
Query Match 5.9%; Score 512.5; DB 4; Length 3169;
Best Local Similarity 20.2%; Pred. No. 4.2e-22;
Matches 422; Conservative 333; Mismatches 720; Indels 619; Gaps 104;
Qy 14 YVKFRI---QDEYVAINAL--BEYHNSSSVVVEKYKLDKINLNDNLYNTYKSGR 68
Db 160 YIKIRKTRGAEDQTTITQSLIINELLNGVDRNTI-PFKISELNDIHSVENMQIKSR 218
Qy 69 N--KALKKFEYLTVEVLELKNSS-----LTPVEKNLHFTWIGQ 106
Db 219 KGEILVKQGEILLSSLLNDKNGKOLSDNASKIINLLGIEYQSHKVDIEPFIHAWVAGA 278
Qy 107 INDTAINYINQKVDNSDYTVKVFYDSNAF-----LINTLTKT----- 144
Db 279 PPDNTFSVITAFNTYKYDTYLLWIDPNAGAAKESGILKNIAAMYAIRLRRTNPHLAE 338
Qy 145 -----IVESATNNTLE--SPRENLDPEFDY-----NKFYRKMEIYYDKQKHF 187
Db 339 ENNEVILAKTQNETIEFKETRELRKLENRYKSLTSETKEKPNVFFLESIMQDNVYF 398
Qy 188 DYKQSQIENPEFI-----IDNIK--TYLSEYKQLEALNKYIEESLNKITANGN-- 238
Db 399 TYCISNGISNTDDISRLDFLTNVLKLSPEVQNDFKSTVEKNKRDIDLLKNTISQFGDRF 458
Qy 239 ---DIRNLEKFADEDLVRLYNQELVERNLAAASDILASMLKEDGGVYLDVDILPGIQP 295
Db 459 QLRDINTLESFKKPPQDYFYQOEMLLRWYAAASQVRAINILKYGIGYITDITDILPAYSD 518
Qy 296 DLFKSINKPDSITNTSWEMIKLEAIMKYK--EYIPG--YTSKNFMDLDEEVQSPESALS 351
Db 519 KVSQIINE-KSDDKPFEDLKLRIISSEILSLIKEKYSIKH-DGLDETTLNQLNLI- 575
Qy 352 SKSDXSEIFPLPLDDIKVSPLEVKIAPANSVI-----NQALISLKSYC 395
Db 576 SEIEK---LTIDDY-FKPVEKVVYRDTFKIYKQYKWTENTNIRGNMFLTKGSKC 630
Qy 396 SLVINGIKRYKILNDNLNPSINEGTDFNTMKFSDKLASISNE-----DNM 444
Db 631 IDFILSGQKKQYLEL-QRIRDNISYNLFTYT-----EDLKSNNVAIGGIPAKKYLEHG 684
Qy 445 MEMIKITNVLKVGAPDVRSTINLSGP----- 471
Db 685 LF---SEYRQDGTTPYVVSTLINISGPDMMQMKYKYSIGRIGEVHKKNLSDVNF 740
Qy 472 GYVTGAYQDLLMF---KONSTMIHLEPE-----LRNFEPPKTKISQLT-----BQEITS 518
Db 741 GYVASSKNKSNFWLNPVSGINDITPDDESSWAVRNDINKILFEKINCHVPEKLPST 800
Qy 519 LMSFNOARAKSQFEYKKGYPFEGALGEDNDLFAQNTVLDKDYVSKKILSSWKTNRKEYI 578
Db 801 LY-----VEIDSRSFQGW---DNKSIKHVTEINKDLI--KQINLLLTSSNIDV 844
Qy 579 HVIYVQLQGDKISYEASCNLFSDKPYSSILYQKNISETAYVYVYVADAIEIKDYRIPY 638
Db 845 KLLIKL--DRELYAISKI-----DNPJALRSIRTQLQLANYVTNTPEPENTINFIY 896
Qy 639 QISNKEN-----IKLTFIGHGKSEFNTDTFANLDVDS-----LSSEIEFT 677
Db 897 DFYRKQDQLLSAUKL-----FRNADTKLIIVYNSVMEKNVFLAEVISC 942
Qy 678 ILMARADISPKYIEINLIGCNMFYSISAEETYPKLLALKDKRVSELMPSISQDSI-- 735

Db 943 VLRSKVD---SYINEN-----KQNLKEDA--GALRDYAKLKMKEFLSMLDDGDK 990
Qy 736 TVGANQYVRINEGKREILDHSGKWKINKEEIIKDIKSEKISFNP-----K 783
Db 991 IITNAY---IKERDL-----SGIINIENSIHGESFDIIRNOHEWGLSTVEQFK 1042
Qy 784 ENKIIVKSKYLHELSTLQEIIRNANSDDIDLEKKVLTCEINVAS-----NIDRQIVE 838
Db 1043 KFEFYKSE--LSSAKSIFDDIKNK-YITDPETKRNVLVHQLSDIKERIAFLDISHYAYP 1100
Qy 839 GRTEAKNLTSDSINKNEFKLIESI-----SDSLYDLKHQNL 878
Db 1101 GSLEKQLS-----GVFSDINIAYLLASVCSGHYGVVPAPSKLLELRHTK 1156
Qy 879 DSHFISFEDISKTFGRIRFINKETGNSI---FIETEKEIFSEYATHISKEISNKT 935
Db 1157 SNEWI--EKITP-----YVYDILSDNVNVRPLSEEQKILNDIKLEISKVS---EQ 1207
Qy 936 IFNVNGKLVKNLDAHEVNTLNSAFTIQLSIEYNTTK-----ESLSNLSV----- 983
Db 1208 YFMKLTQKSSVIGIKYSVDFDYENLFLSLPINQNLTPFMYRFEMLYDIHIGIEN 1267
Qy 984 -AMKVQVYACLFTGLNTIDASKVVELV-----STAL-DETIDL 1021
Db 1268 KAREFTYKFPSSNLDFLINDERVNLNLEGLIKKYKLSLSEIHRLTNSTSPADISPL 1327
Qy 1022 LPTLSELPIIATII-----DGVSLGAIAK-----ELSETNDPLLR 1057
Db 1328 LQII---CPSEITIIKTEYVGHQLTNMTVASVVKPYDFSNLGAINDSKSVDPALH 1384
Qy 1058 QEIE-AKIGMAVNLTAASAIVTSAL-----GIASGPSILL 1093
Db 1385 TIVEQAKYNLLSWNDYFNTHASITWDTARQHKSTNIEFHPQSLDFDRDQKGLGLSLY 1444
Qy 1094 VPLAGISAGIPLVNN-----ELILODKATKVIDYFKHISLAETEGATILDXKIMPQD 1149
Db 1445 LDTGGYGGYQKLRHNTDASTLYOTK-----YNDNLKLSNRDDPFELRTORITMSNE 1498
Qy 1150 LVLSEIDFNNSITLGCETWRAEGSGHTLDDI--DHFFSPSITYRKPLMSIYDVNL 1207
Db 1499 L-----GNRLKNAQLEVK---DPILTGLIYORRISLLITTEYHSLAQOISS 1548
Qy 1208 IKK-----EKIDFSKOLMVLNPNRPFVGYEMGWTFGRSLNDGDKJLD 1252
Db 1549 FWRVTDNPFHGDHPSLAQALTIKNI-----TSNR-----NFSLYSGSI--- 1589
Qy 1253 RIRDPH-EGOFYRYFAFIADALITKUPRYEDTNVRINDG---NTRSPFVPIVTEOI 1308
Db 1590 -VKIYFSESLLNMWKYIKLPLVQTGSLRLDYLTTPKELSTSGGSLNMGHLFV----- 1642
Qy 1309 RKNLSYF-YCGSGYSLSLSPYNNMIDNLVENDTWIDVNVKNITIESDEIQ--- 1363
Db 1643 ---SFIYDIGV-----INGNRIBEST---DVKNKIRSLKINGDIOHYIN 1682
Qy 1364 ---KGELIENILSKLNTEDNKI-----ILN---NHTINFY- 1392
Db 1583 THYLSSEQTQKIDIVDFLQDNTIKVKLESIDKPISEIQPLHSLISLQKHEVKNLLS 1742
Qy 1393 GDINE-SNRF-----ISLTFSEILEDINI-----IIEI---DLVSKSYKILLSGNMKL 1436
Db 1743 GLLDFFSNKLRKQGLSLKTNVLSVNNFKESKINDTVEVTVDLQGRLYRVDITRVIGL 1802
Qy 1437 I--ENSSDIQKIDHIFNGEHQKPIYPIYIDNETKNGFIDYSK-----KEGLFTABFSN 1490
Db 1803 TFKEGINSLSALEHNM-----IDAIMSVIGLVQYARMTKMDNISADHAG 1849
Qy 1491 B-SIIRNIMYDPSNNLFIYSSKDKIIRINKGVK-LLIGNYFKDDMKVLSLFTIEDTN 1548
Db 1850 AVSDIKNIIV-----DKFLGGILITTNRNVNPGVSGASLEGFISSGLEVCAAS----- 1897
Qy 1549 TIKLNGV---YLDENGVAQILKFMNNAKSALNTSNSLNFLESINIKIENYNNLD----- 1600
Db 1898 --RMGGTAGRYL--SNVAKVTKL-----PLLDI--GLINWSLYDSSINLHAKAT 1939

Qy 1601 PNIEFI---LDTNF-IISGNSIGQPE-----LICDCKNIQPIYFINKIKETSVTLYV-- 1650
Db 1940 TQBEYITSTADIVFSSINTALSICATAYPELAIA-----IVPITI-FSHEVKYAVYVNO 1993
Qy 1651 -GNRQNLIVFSPVHLDD-----SGNISSTVINFSQKLYG---IDRYVNVKVI 1694
Db 1994 INERHKLWLAERKVLNDSAKVLSINKATGIIDLSNNQVGLNYLDRENPPI 2047

RESULT 14

US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 4.0%; Score 347; DB 4; Length 10182;
Best Local Similarity 19.4%; Pred. No. 2.5e-11;
Matches 354; Conservative 318; Mismatches 662; Indels 494; Gaps 97;

Qy 20 QDEYVAILNALEYHNMSSSVVEKYLKLDINN-----LTDN-----YLN-- 61
Db 7557 QKD---TILNHI---FSAPTRSQVGEKIASAKQLNNTMKALRDSIADNNEILQSSKYFNE 7611
Qy 62 TYKSGRKNALKKPEYLTWELKKNLSLTPV-----NDQPTFVMADEIOSVLNEVKQTKDNLH-- 105
Db 7612 SEQQNAVYQNAKAKNII-----NDQPTFVMADEIOSVLNEVKQTKDNLH--GD 7659
Qy 106 Q--IND-----TAINVINQWKDVNSDYTVKYFYDSN-----AFLINTLKKT 144
Db 7660 QKLANDKTDQAATNALNLYNQAGNLETQVQ---NSNSRPEYQVKVQLANQLNDAMKK 7716
Qy 145 IVESATNNLTSPRENLDPEFDYKFKRMEIYDKQKHFDIYYSQIEE--NPEF-- 200
Db 7717 LDDALTGN--DAIKQTSNVINETSQ-----QVNFDE---YTDGRKNIVAEQTPNNSP 7765
Qy 201 -----IIDNIIKTVLSNEYSKDLEALNKY-----IBESLNKITANNNGDIRNLEKFADEDLV 252
Db 7766 TNINTIADKI-----TEAKNDLHGQVKLQAOQSINTI-----NQMTGLNQAKQEL- 7813
Qy 253 RLYNQELVERWNLAASDILRISMLKEDGGVYLDVLDILPGIQPLFKSINKPDSITNTSW 312
Db 7814 ---NOEIQOQOTRSEVHGVINKAQAALNDS-----MNTLRQSTIDEHVKQTSN 7858
Qy 313 EMIKL-----EAIMKYKEIYIPGTSKNPDMLEEVQORSFESALSCKSKSIFLPLD 364
Db 7859 YINETVGNQATYNNNAVDKVKIINQTSNPTNPL--EVERATSNVKLSK-DALHGERELN 7915
Qy 365 DIKVSPLVEKAFANNSV--INQALISLKDYSYCDL-----VINQIKNRYKIIN----- 411
Db 7916 DNKNSK-----TFAVNHLDNLQA---QKEALTHEIEQATIVSQVNNIYNKAKALNNDMK 7967
Qy 412 -----DNLNPSINEGTFTNTMKLFPDKLASISNEDNMFWI-----KITNYL 454
Db 7968 KLDKIVAQODNVQRQSNMYNINEDSTPQNMVND---TINHAQSIIDQVANTPTWSDHEIENAI 8024
Qy 455 KYGFAPDVRSTIN--LSGPGVVTGAVQDILLMFKDNSTNIHLEPELRNPEFF-KTKISQLT 512

Db 1307 EWKINSQ--SDPRYQLTKNCSSIVARILKAGGADQLIGHHMKPRGIWTPNELYKFSQ 1364
QY 405 -----NRYKILNMLNPSINEGDTFNTTKIFSDK-L 435
Db 1365 KIQEARLAQIAVOQCKPIHNSLOALSDHNKVAINDGTSF--NERKNLSPLTRFFNDHFF 1422
QY 436 ASISNEDNMFMKITYNLKVGFPADV-RST----- 465
Db 1423 GSYERRDMMTVMRKNEKITLKGDAGRUTGSYRGNDNIPBATTCKVVLFLHGSNSPT 1482
QY 466 -----INLSGPGVYTGAYQDILLMFKDNSTNIHL-----EPE-- 497
Db 1483 EKQSSSFHYNNQOGIDMLAINRFGESDGSFTEQWYADATMPRYLVNDKIDPKNI 1542
QY 498 -LRNFEPKTKISOLTE-----QEITSLW-----SFNQAPAKSOFPEYKGYFEGALGED 546
Db 1543 IHHGYSNGAPITAAKIASDISANGQHIALFLDRPMSMSKAIESYDEYSLKLTQLAKK 1602
QY 547 DNLDFP-----QNTVLDKDYVSK-----KILSSM 570
Db 1603 INGQFSVEKNLQGSFKDIPILLTDCGELGIGGEKMKTKLLDKGYLVKGERTDVSHLSL 1662
QY 571 KTRNK-----EYIH----- 579
Db 1663 KLMQYQKQISSLSPSEYVNTNHRNALDSIONPIKKNKVTGWVYVKNPNDGDSRF 1722
QY 580 ---YIVOLQGDKISYEASCLFSDPYSSILYQKNEGSETAYYYYVADAEIKEIDKYRI 636
Db 1723 DSQIIQTEDDPIVAQASALAAKAKNSIVQLDADG-----HYRVA-----YGD 1768
QY 637 PYQISNKRNIKLTPIGHGK--SEFNTDTFANLDVDSLSSEI---ETILNLA-KADISPKY 690
Db 1769 PAQLSGK--IRWLQVCHGRETTENNHLRSLNNADELSQLVKFNTVFSKNNINTPEH 1826
QY 691 IEINLGCNMFYSISABEYTPGKLLIKIDRVSELMPSISQDSI--TVSANGYEVRIE 748
Db 1827 ISI--VGCISLNDKQAGFAH-----QFINALDQOQIRSSVSARVTEVAVDS 1871
QY 749 EGKEILLDHSKWKINKEESIIOKISSKEYISFNPKENKIIVKSKYLHELSTLLQETRNNA 808
Db 1872 NGHYTKDNNGEWAKO-----NQKVVNNKKEGKIITEFEQIUNGV 1914
QY 809 NSSDIDLEK-----KVMLTECEINVASNIDQIVGRIEAEKNTSDSINYKN--- 857
Db 1915 AEGDIDLTKIGASADNKVRGAIAADNNEVFTAPKKGKSTKQSSSGSDNSLSYSGNIQVS 1974
QY 858 -----EFKLIESISDSL-----YDLKHQGLDSDSHFISFED 888
Db 1975 VGDGEFTVNWGTSNLGKIKVGTGFKSILVFGDNNVMVHVGNGDSKH--SVDIAGYQAFEG 2032
QY 889 ISKTENGFRIPFPINKETGNSIFITEKEIFS-----EYATHISKEISNIKDTIFDNVN 941
Db 2033 MQLFVGTRNVSF-NLGRSNDLIVMLEKSIPTPPLINFPDGAARITKYLESIACSGENNDN 2091
QY 942 GKL-----VKKVNLDAAEVNTLNSAPFIQSILEVN-----TTKESLSNLSVAM 985
Db 2092 DWLSVQNDQWTLGAKKFAIDNSGIDQTSNDVY--KTLTDLNSQNKSSRGLKSDLEATL 2149
QY 986 KVQVYAQLFSTGLAT---ITDASKVVELVS-TALDETIIDLLPTLSEGLPIIATIID-GV 1039
Db 2150 NKKYNQWLGSGNNEMGKISRAEQRLQDLSKIAFNFAIG-----GQADIQVTAGNWNL 2204
QY 1040 SLGAIAKELSTNDPLURQETAEKIGIMAVNLTA---ASTAIVTSALGIASGF-SILLVP 1095
Db 2205 MFGDNIQSIIDTN-----LGSFLGLMTQOFTATGMVETFTYTPQNLPKLKKNLLGR 2257
QY 1096 LAGISA-----GIPSLVNNEL-----ILQD-----KAT 1118
Db 2258 LASVSSDRTLGBIFGVDSTTGKTIISRNGEPIDVVAIKOMLAVVTEFGGEKLSAFTDPT 2317
QY 1119 KVIVDYFKH-----ISLAETGAFLLDDKIMPODDLVLSEIDFN-NNSITLGKCE 1168
Db 2318 KLLDNLQSSINLKGDIISFAESHG-----LXEK--ATEDQQNESEVSINAQTPITTSQTE 2371

QY 1169 IWRAEGSGSGHLTDDIDHFRSSPSITYRKPMLSIV-----DVLN 1207
Db 2372 -----NSKITEENDRPPGFNSLNPFPATIFNKDKQOTDMRDLAENLKENLAADVLN 2423
QY 1208 IKKEKIDFSK-----DLMWLPNAPNRVFGYEMGWTGFRSLDNDGTKLLDRIRDHY 1258
Db 2424 MEKTLDFLRNSGHLKGDGDHIV-----SLGNYNFNW-----GGDNDLGAYLGDN 2470
QY 1259 EGQFW-----RVFAFIADALIT-----KUKPRYEDTNVRINLDGNTSRFIVPVITTEQ 1307
Db 2471 N---FWGGRGSDTFVATGISNIFTGGYNDLGVLGRSNMFMFGGDGDDTALIAGRINNVY 2527
QY 1308 IRKNLSYSF-YGSGG-----SYSLSPYNNIDNLNVLNENDTWIDVNVVKNITIE 1358
Db 2528 LGDGLDKAFVFGEGGEIHTNAGNDYAVTTGNYNRFSGS--EQD-FVVTIGN-HNQISLE 2583
QY 1359 -----SDEIQ-----KGELIENILSKL-----NIEDN 1380
Db 2584 BONDPAKIFGNVNRNLNGGGGDEIQLMGYHAVVGGDNDQJIAASFSKFSLLNGTND 2643
QY 1381 KIILNNHTINFYGDINESNRFIISLTFSILEDINIIE--IDLVSYSKILLSG-NCMKLI 1437
Db 2644 IILGGYQNHFKGGNGIDSPJLN-----NNVIDCYIDDISQEDNIVLGDIDWNKLM 2694
QY 1438 ENSDIOQKIDHIGPNGEHQKIPIYSYIDNETKINGFIDY 1477
Db 2695 PERSGYDLKISHIRPPQNTGEATFERIGSTT-----FTDY 2730

Search completed: April 1, 2004, 16:46:06
Job time : 35.625 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 16:33:50 ; Search time 71.25 seconds

(without alignments)
6741.484 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1700

Perfect score: 8675

Sequence: 1 MVLVNAQLQKVVYKFRIQ.....VLYGIDRYVKNVLIAPNLYT 1700

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6756.5	77.9	2366	2 AAR95011	Aar95011 C. diffi
2	6756.5	77.9	2366	2 AAW68388	Aaw68388 Clostridi
3	4467	51.5	2710	2 AAR95016	Aar95016 C. diffi
4	4467	51.5	2710	2 AAW68387	Aaw68387 Clostridi
5	2792	32.2	546	2 AAY33700	Aay33700 C. sordel
6	512.5	5.9	3169	6 ABO14444	Abol14444 E. coli 0
7	431.5	5.0	4688	6 ABU48941	Abu48941 Protein e
8	398	4.6	1979	3 AAB18171	Aab18171 Plasmodiu
9	374.5	4.3	3043	6 ABM68924	Abm68924 Photorhab
10	372	4.3	2410	6 ABM68965	Abm68965 Photorhab
11	355.5	4.1	5024	4 AAG82935	Aag82935 S. epider
12	347	4.0	6641	6 ABU42656	Abu42656 Protein e
13	347	4.0	10182	5 ABP38314	Abp38314 Staphyloc
14	343	4.0	2485	3 AAB18172	Aab18172 Plasmodiu
15	340.5	3.9	5005	6 ABU48947	Abu48947 Protein e
16	336	3.9	4620	6 ABU40997	Abu40997 Protein e
17	313.5	3.6	3973	3 AAB18253	Aab18253 Plasmodiu
18	312	3.6	1639	2 AAW54145	Aaw54145 P. falcip
19	312	3.6	1639	5 AAE29345	Aae29345 Plasmodiu
20	311	3.6	1516	3 AAB18195	Aab18195 Plasmodiu
21	305.5	3.5	2496	3 AAB18222	Aab18222 Plasmodiu
22	304.5	3.5	1802	3 AAB18217	Aab18217 Plasmodiu
23	304	3.5	2184	4 AAE00425	Aae00425 P. falcip
24	302.5	3.5	1254	2 AAR07503	Aar07503 Merozoite
25	302.5	3.5	1254	2 AAW24575	Aaw24575 Merozoite

26	302.5	3.5	1370	6 ABM71130	Abm71130 Staphyloc
27	298.5	3.4	2539	3 AAB18198	Aab18198 Plasmodiu
28	296.5	3.4	1308	3 AAB18167	Aab18167 Plasmodiu
29	296	3.4	1654	1 AAP50777	Aap50777 Sequence
30	293.5	3.4	1844	3 AAB18250	Aab18250 Plasmodiu
31	291	3.4	2295	3 AAB18180	Aab18180 Plasmodiu
32	290	3.3	1182	6 ABU25521	Abu25521 Protein e
33	289	3.3	2575	6 ABU44210	Abu44210 Protein e
34	289	3.3	3433	2 AAW22017	Aaw22017 Utraphin.
35	288	3.3	1875	6 ABR53560	AbR53560 Protein s
36	288	3.3	2000	6 ABR52698	AbR52698 Protein s
37	287	3.3	1610	7 ADB46133	Adb46133 Plasmodiu
38	286.5	3.3	2046	6 ABU44208	Abu44208 Protein e
39	286.5	3.3	2681	6 ABJ19025	Abj19025 Pathogen
40	285.5	3.3	1948	5 ABP73774	Abp73774 Candida a
41	285	3.3	2010	3 AAB18218	Aab18218 Plasmodiu
42	283.5	3.3	1881	5 ABP73809	Abp73809 Candida a
43	281.5	3.2	1196	6 ABU24813	Abu24813 Protein e
44	281	3.2	3692	6 ABU43311	Abu43311 Protein e
45	280	3.2	1436	3 AAB18199	Aab18199 Plasmodiu

ALIGNMENTS

RESULT 1

AAR95011
ID AAR95011 standard; protein; 2366 AA.
XX
AC AAR95011;
XX
DT 16-OCT-2003 (revised)
DT 08-JUL-1996 (first entry)
XX
DE C. difficile toxin B.
XX
KW Toxin B; cytotoxin; enterotoxin; fusion protein; antitoxin; diarrhoea;
KW therapy; diagnosis; vaccine.
XX
OS Clostridium difficile; VPI strain 10463 (ATCC 10463).
XX
PN WO9612802-A1.
XX
PD 02-MAY-1996.
XX
PF 23-OCT-1995; 95WO-US013737.
XX
PR 24-OCT-1994; 94US-00329154.
PR 16-MAR-1995; 95US-00405496.
PR 14-APR-1995; 95US-00422711.
PR 07-JUN-1995; 95US-00480604.
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Padhye NV, Kink JA, Thalley BS, Stafford DC;
PI Firca JR;
XX
DR WPI: 1996-230603/23.
DR N-PSDB; AAR29247.
XX
PT Fusion proteins comprising non-toxin protein and part of toxin - useful
PT to form anti-toxins against Clostridium botulinum type A, and C.
PT difficult type toxins, and to treat C. difficile intoxication, partic.
PT diarrhoea.
XX
PS Claim 36; Page 313-323; 434pp; English.
XX
CC Clostridium difficile VPI strain 10463 toxin B (AAR95011), the product of
CC the toxin B gene (AAR29247), is a cytotoxin associated with diarrhoeic
CC disease. It can be obtd. by expression in transformed E. coli hosts of
CC portions of DNA that together cover the entire toxin B gene. Toxin B, and
CC portions of it (see also AAR95012-13, AAR75371-72 and AAR95018), pref.
CC expressed as fusions to polyhistidine affinity tags or maltose binding

CC protein, are used to raise avian antibodies useful as antitoxins or
CC diagnostics, and in vaccine prodn. (Updated on 16-OCT-2003 to standardise
CC OS field)

SQ Sequence 2366 AA;

Query Match	77.9%;	Score 6756.5;	DB 2;	Length 2366;
Best Local Similarity	76.5%;	Pred. No. 0;		
Matches 1301;	Conservative 192;	Mismatches 206;	Indels 1;	Gaps 1;

Qy	1	MNLVKAQLOKMYVKFRIQDEYVALINALAEYHNHNSSSVVEKYLKLOINNLITDNYL	60
Db	1	MSLVNRKQLEKMANVRFQDEYVALDLAEYHNHNSENTVVEKYLKLOINSLTDIYI	60
Qy	61	NTYKSGRNKALKKFEYLTWEVLKKNLSLTPVEKNLHFITWIGGQINDTAINYNOWKD	120
Db	61	DTYKSGRNKALKKFEYLVTEVLKKNLTPVEKNLHFVWIGGQINDTAINYNOWKD	120
Qy	121	VNSDYTVKYFYDSNAFLNTLKKTIVBSATNNTLESFRENLDPEFDYKFKRMEIYY	180
Db	121	VNSDYNNVAFYDSNAFLNTLKKTIVBSAINDTLESFRENLDPRFDYKFKRMEIYY	180
Qy	181	DKQKHFIDYKGOIBENPEFIIDNIIKTYISNEVSKOLEBALNKYIBESLNKTANNXNDI	240
Db	181	DKQKHFINYKAQREENPELIIDDIIVKTYISNEYSKEIDELNTYIBESLNKTQNSGNDV	240
Qy	241	RNLKXFADEDLVRLYNQELVBRWNLAAASDILIRSMLEKEDGVYLDVDILPGIQDLPKS	300
Db	241	RNPEBFGKSGENLVEQELVBRWNLAAASDLIRLSALKEIGGYLDVDMPLGQDLPES	300
Qy	301	INKPDSITTWSEMIKLEBAIMKYKEYIPGYTSKNFMDLDEEVQSFESALSXSDESEIF	360
Db	301	IEKPSSTVDFWEMTKLEBALMKYKEYIPEYTSHEFMDLDEEVQSFESVLAKSDKSEIF	360
Qy	361	LPDDEIKVSPLEVKITAPANNVINOALISLXDSVCSDLVINTQIKRYKIINLNLPSTNE	420
Db	361	SSLGDMEASPLEVKIAFNASKGIINQGLTSVKDSCSNLIVKQIENRYKILNNSLPAISE	420
Qy	421	GTDENNTMKIFSDKLASISNEDNMWMIKINYLKVGFPADVSTNLGSPGVYTGAYOD	480
Db	421	DNDENFTNTWTFIDSIMAFANADNGRFMELGKYLURGVFPDVKTTNLSGSPRAYAAAYOD	480
Qy	481	LMWFKDNSTNIHLLPELERNPEFPKTKISQUTEQITSLMSFNQARAKSQFBEYKKGYPE	540
Db	481	LMWFKEGSNHNLIEADLRNFEISKTNISQSTEQEAMSLASFDARAKAQFBEYKENYPE	540
Qy	541	GALGEDNDLDFQANTVLOKOVYVSKILGSMKTRNKYIHYIVLOQDKKISYASACNLPSK	600
Db	541	GSUGEDNDLDFSQNTIVDPKVELLEKISSLASRBSRGYIHYIVLOQDKKISYEAACNLFAK	600
Qy	601	DPVSSILYKNTIEGSETAYYYVADABETKEDKYRIPQJSNKENTKLFPIGHGKSEFT	660
Db	601	TPVDSVLFOKNEDESEIAYYNGPGEIQEIDKYPISIIISDRPKILFTPIGHGKDEFNT	660
Qy	661	DTFANDVDLSLSEETTINLAKADISPKYTEIINLLGCNMFSYSISAEETYPGKLLKIK	720
Db	661	DIFAGFDVDSLSTEIEAADLAKEDISPKYTEIINLLGCNMFSYSINVEETYPGKLLKVK	720
Qy	721	DRVSELMPSISQDSITVSAQVEVRAINBEGKREILDHSGKWINKBESSIKOISKEYISF	780
Db	721	DKISELMPSISQDSITVSAQVEVRAINBEGRELLDHSGEWINKBESSIKOISKEYISF	780
Qy	781	NPKENKIIVKSXYLHELSTLQETIANNANSDDILEKKWMLTECEINVASNIDROIIVER	840
Db	781	NPKENKIVKSKNPLSTLQETIRNNNSDDILEBKWMLTECEINVASNIDTQIIVER	840
Qy	841	IEBAKNLTSDSINYIKNPFKLTIESDLYDLKHQNGLDDSHTSFEDISKTENGPFRIF	900
Db	841	IEBAKNLTSDSINYIKDPEKLTIESDLDLCKQONELEDSHTSFEDISDETSDEGSPFR	900
Qy	901	INKETGNSIFETEKEIFSEYATHISKISNIKOTIFDNVNGKLVKKNVLDAAHEVNTJNL	960
Db	901	INKETGESIFVETEKITFSEYANHITBESIKIGTIFDVTNGKLVKKNVLDTHEVNTJNL	960

Qy	961	SAPFIQSLIYNTTKESLSNLSVAMKVQVYAOI	FSTGLNTIITDASKVVELVGTALDETID	1020
Db	961	AAAFIQLIYNSKESLSNLSVAMKVQVYAOI	FSTGLNTIITDAKVKVELVGTALDETID	1020
Qy	1021	LIPTLSEGLPIITATIIDGVSGLGAAIKELSETNDPLL	QBIIEAKIGIMAVNLTAASATVIT	1080
Db	1021	LLPTLSEGLPIITATIIDGVSGLGAAIKELSETDPLL	QBIIEAKIGIMAVNLTTATTAIT	1080
Qy	1081	SALGIASGFSIILVPLAGISAGIPSLVNNELIQDKATK	VIDYFKHISLAETEGAFITLID	1140
Db	1081	SSIGIASGFSIILVPLAGISAGIPSLVNNELVLRKAT	KVVDYFKHVSIVETEGVFTLID	1140
Qy	1141	DKIIMPQDVLVSEIDFNPNNSITLGKCEIWRAGSGHTL	TDIDIHFFSPSITYRKPMIL	1200
Db	1141	DKIIMPQDVLVSEIDFNPNNSIVLGKCEIWRMEGGSGH	TVDIDIHFFSAPSITYREPHL	1200
Qy	1201	SIYDVNLINIKKEIDPFKDLMLVLPNAPNRVPGVEMGH	TGFRSLDNDGTKLDRIDHYEG	1260
Db	1201	SIYDVLEVEKEELDLSKDLMLVLPNAPNRVPAWETGT	PGRLSLENDGTKLDRIDRINYEG	1260
Qy	1261	QFYWRVFAPTADALITKLPRVEDTNVRINLQDNTSFIV	PVTITBOIRKNSLSYSFYGSG	1320
Db	1261	EFYWRVFAPTADALITTLKPRVEDTNIRINLQDNTSRF	IVPITITBYIEKLSYSFYGSG	1320
Qy	1321	GSYSLSLSPYNNMIDLNVENDTWIIVDNVVKNTIETSE	DIQKGBLIENILSKMLIEDN	1380
Db	1321	GTVALSLSQYNWGINTELSESDVWIIDVNVVEDVTIE	SDKIKKGBLIEGILSTLSEEN	1380
Qy	1381	KIILNHRITNFYGDINESRPFISLTSLIEDINIIIEID	LVSKYSKILSGNCMKLIENS	1440
Db	1381	KIILNSHEINFSGVNGSGFVSLTSLIEGINAIIEDVL	LSKYSKILLISGELKILMNS	1440
Qy	1441	SDTQOKIDHIGFNGEHQKIPIYSIINETKYNGFIDYS	KKEGHFTAFESNESIIRNIYMP	1500
Db	1441	NHIQOKIDYIGFNSLQKNIPIYSFVDSBKGXNGFING	STKEGHLFVSELPDWLISKVYMD	1500
Qy	1501	DSNNLFIYSSKDJLDRIIRIINGDVKLLIGNYFKDD	MKVSLSFTIEDNTITKLNGVYLDE	1560
Db	1501	DSKPSFGYSGNNLQKVKITKQNVNLTGYLLKDDIKI	LSLSTLQDEKTIKUNSVHLDES	1560
Qy	1561	GVAQILKFNNAKSALNTSNLNFLESINIKFIYNNLD	PNPTIEFIDTNTFIISGNSIG	1620
Db	1561	GVAEILKFMR-KONTNITSLSLMSFLESNMKISIFVN	FLQSNIKFILDANFIISGTTISG	1619
Qy	1621	QFELICDKXNIQPYFINFKIETSUTLYVGNFQNIIV	BPESYXHLDSGNSISSTVINFSQK	1680
Db	1620	QFEPICDENDNIQPYF-KPNTLETNTVLYVGNFQNNI	VEPNYDLDSGDISSTVINFSQK	1679
Qy	1681	YLYGIDRYVKNVIAAPNLYT	1700	
Db	1680	YLYGIDSCVNKWIAPNLYT	1699	

RESULT 2
AAW68388
ID AAW68388 standard; protein; 2366 AA.

AA
AC
AAW68388;

07-DEC-1998 (first entry)

XX DE Clostridium difficile toxin B.

XX Antitoxin; vaccine; cytotoxin; toxin B; intoxication; immunogen;
KW pseudomembranous enterocolitis.

OS Clostridium difficile.

PN W09808540-A1.

05-MAR-1998.

PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
XX WPI; 1998-230234/20.
DR N-FSDB; AAV30561.
DR
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
XX
FS Example 18; Page 241-249; 429pp; English.
XX
XX This is the amino acid sequence of Clostridium difficile toxin B, deduced
CC from the coding region (see AAV30561) of the toxin B gene. Fragments of
CC the toxin B gene have been cloned into various prokaryotic expression
CC systems, and assessed for the ability to express recombinant toxin B
CC protein in E. coli. It would be advantageous to use simple and
CC inexpensive prokaryotic expression systems to produce and purify high
CC levels of recombinant toxin B for immunisation purposes. The invention
CC specifically relates to recombinant proteins derived from Clostridium
CC botulinum toxins (see AAV8389-400) and their use as immunogens for the
CC production of vaccines and antitoxins
XX
SQ Sequence 2366 AA;

Query Match 77.9%; Score 6756.5; DB 2; Length 2366;
Best Local Similarity 76.5%; Pred. No. 0;
Matches 1301; Conservative 192; Mismatches 206; Indels 1; Gaps 1;

QY 1 MNLVYKALQKQVYVKKRIEODEVATLNALEEHYHNSSESVVEKYLKDKINLNDYIL 60
DB 1 MSLVNRKQLEKMANVRFTDEDEVALDLEAEYHNSENTVVEKYLKDKINSLDVI 60

QY 61 NTKKSGRNKALKKFKELYLMEVLEKNNSLTPVEKNLHPFIWGGQINDTAIYINQWKD 120
DB 61 DTKKSGRNKALKKFKELYLMEVLEKNNSLTPVEKNLHPFIWGGQINDTAIYINQWKD 120

QY 121 VNSDYTVKVDNSAFNLITLTKTIVESATNNTLESRENLDPEPDYKPKRMELIY 180
DB 121 VNSDYNVNFVDSNAFLITLTKTIVESAINDTLESRENLDPEPDYKPKRMQIY 180

QY 181 DKQKFTDYKSOLEENPERIIONIKTYLSNEVSKDLEALKNYIEESLNKIITANGNDI 240
DB 181 DKQKFNINYKAQREENPELIIDIVKTYLSNEYSKEIDELNTYIEESLNKIITANGNDV 240

QY 241 RNLEKFADEDLVRLYNQELVERMNLAAASDILRISMLKEDGGVYLDVILPGIQLPKS 300
DB 241 RNLEKFADEDLVRLYNQELVERMNLAAASDILRISALKEGGVYLDVILPGIQLPKS 300

QY 301 INKPDSTNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVORSPESALSSKSKSEIF 360
DB 301 IEPSSVTVDFWEMTKLEAIMKYKEYIPEVTSEHFDMLDEEVORSPESVSLASKSKSEIF 360

QY 361 LPDDIKVPLEVKIAFNANSVINQALISLKDSCDLVNTQKNRYKILNDNLNPSINE 420
DB 361 SSLGMEASPLEVKIAFNANSVINQALISLKDSCDLVNTQKNRYKILNDNLNPSINE 420

QY 421 GTDENTMTKIPSDKLASINEDANMFMKITYLKVGPAPDVASTNLSGPGVYTCAYOD 480
DB 421 DNDENTMTKIPSDKLASINEDANMFMKITYLKVGPAPDVASTNLSGPGVYTCAYOD 480

QY 481 LLMEKDNSTNHLPELNRPEPKTKISOLTEOITSLWSFNOARAKSOFEEYKKGYPE 540
DB 481 LLMEKDSMNHLPELNRPEPKTKISOLTEOITSLWSFNOARAKSOFEEYKKGYPE 540

QY 541 GALGEDNDLPAQNTVLDKDYVSKKILSSMKTRNKYIHYIVQLQGDKISYEASCNLFK 600
DB 541 GSLGEDNDLDFSNVVDKDYVSKKILSSMKTRNKYIHYIVQLQGDKISYEACNLFAK 600

QY 601 DPYSSILYQKNIEGSETAYVYVADAEIKEDKYRIPYQISNKNKIKLTFIGHGKSEPT 660
DB 601 TPYDSVLFQKNIEDEIAYVYVADAEIKEDKYRIPYQISNKNKIKLTFIGHGKDEPT 660

QY 661 DTFANLDVDSLSSEIETILNLAKADISPKYIEINLLGCMFSPYSISABETYPGKLLKIK 720
DB 661 DIFAGFDVDSLSTEIEAAILAKEDISPKSIEINLLGCMFSPYSINVEETYPGKLLKVK 720

QY 721 DRVSELMPSISQDSITVSAQYEVRIINDEGKREILDHSGKWKINKEESIIDKISKEYISF 780
DB 721 DKISELMPSISQDSIIVSAQYEVRIINDEGKREILDHSGKWKINKEESIIDKISKEYISF 780

QY 781 NPXENKIIVKSKYHELSTLQEIERNANSDDILEKKVWLTECEINVASNIDQIVEGR 840
DB 781 NPXENKIIVKSKYHELSTLQEIERNANSDDILEKKVWLTECEINVASNIDQIVEGR 840

QY 841 IEAANKLTSDSINIKNEFKLIESISDSLYDLKQNGLDDSHFISFEDISKTENGFRIRF 900
DB 841 IEAANKLTSDSINIKNEFKLIESISDSLYDLKQNGLDDSHFISFEDISKTENGFRIRF 900

QY 901 INKETGNSIFITEKEIPSEYATHISKEISNIKOTIFDNVNGKLVKKNLDAAEVNTLN 960
DB 901 INKETGESIFVETEKTIPEYANHITEBISKIKGTIFDTVNGKLVKKNLDTTHEVNTLN 960

QY 961 SAFFIQSILIEYNTTKESLSNLSVAKVQVYLAQLFSTGLNTITDASKVVELYSTALDETID 1020
DB 961 AAFPIQSILIEYNTTKESLSNLSVAKVQVYLAQLFSTGLNTITDASKVVELYSTALDETID 1020

QY 1021 LLPTLSEGLPIIATIDGVSIGAAKELSETNDPLRQEIKAIGIMAVNLTAATASTAVT 1080
DB 1021 LLPTLSEGLPIIATIDGVSIGAAKELSETNDPLRQEIKAIGIMAVNLTAATASTAVT 1080

QY 1081 SALGIASGFSILLVPLAGISAGIPSLVNNELILOKATKVIDYFKHISLAETEGATLLD 1140
DB 1081 SSLGIASGFSILLVPLAGISAGIPSLVNNELILOKATKVIDYFKHISLAETEGATLLD 1140

QY 1141 DKIMPQDVLSEIDFNNSITLKGCEIWPAGEGSGHTLDDIDHFFSPSPSYRKPWL 1200
DB 1141 DKIMPQDVLSEIDFNNSITLKGCEIWPAGEGSGHTLDDIDHFFSPSPSYRKPWL 1200

QY 1201 SIYDVNTAKKEIDFKDLVLPNAPRVFGEYVGTGPRSLDNDGTCKLLDRDRHYEG 1260
DB 1201 SIYDVNTAKKEIDFKDLVLPNAPRVFGEYVGTGPRSLDNDGTCKLLDRDRHYEG 1260

QY 1261 OFYWRYPAFIADALITLKPRIEDTNVRINDGNTRGFIVPVIITEQIRKLSYSFYGSG 1320
DB 1261 OFYWRYPAFIADALITLKPRIEDTNVRINDGNTRGFIVPVIITEQIRKLSYSFYGSG 1320

QY 1321 GSYLSLSPYNNIDNLVENDTWIDVDNVYKNITTESDEIQKGELIENILSKNTEEN 1380
DB 1321 GSYLSLSPYNNIDNLVENDTWIDVDNVYKNITTESDEIQKGELIENILSKNTEEN 1380

QY 1381 KIILNNHTINPYGOINESNRFSITFSEILEDINIIEIDLVSQYKILLSONCMKLTENS 1440
DB 1381 KIILNNHTINPYGOINESNRFSITFSEILEDINIIEIDLVSQYKILLSONCMKLTENS 1440

QY 1441 SDIQOQIDHIGFNGEHQKYPYSYDNETKNGFIDYSKKEGLTAEFSNESIIRNYMP 1500
DB 1441 SDIQOQIDHIGFNGEHQKYPYSYDNETKNGFIDYSKKEGLTAEFSNESIIRNYMP 1500

QY 1501 DSNLNFYSSKDLKDIRIINKGDVKLLTIGNYFKDKMVSFTTIEDTNTIKNGVYLDEN 1560
DB 1501 DSNLNFYSSKDLKDIRIINKGDVKLLTIGNYFKDKMVSFTTIEDTNTIKNGVYLDEN 1560

QY 1561 GVAQYALKFPMNAKSAINTNSLMPFLESINIKNIFYNNLDPNIBFLDTNFIISGNSIG 1620
DB 1561 GVAQYALKFPMNAKSAINTNSLMPFLESINIKNIFYNNLDPNIBFLDTNFIISGNSIG 1620

QY 1621 QPELICDKDKNIQYFFINFKIKETSITLVGNRQNLIVEPSYHLLDDSGNISSTVINFSQK 1680
DB 1621 QPELICDKDKNIQYFFINFKIKETSITLVGNRQNLIVEPSYHLLDDSGNISSTVINFSQK 1680

QY 1680 QPELICDKDKNIQYFFINFKIKETSITLVGNRQNLIVEPSYHLLDDSGNISSTVINFSQK 1680
DB 1680 QPELICDKDKNIQYFFINFKIKETSITLVGNRQNLIVEPSYHLLDDSGNISSTVINFSQK 1680

QY	1681	YLYGIDRVYKVIAPNLYT	1700	Db	120	INAEYNIKLYWDSYEAFLVNTLKAIVESSTTEALQLLEEEIQNPQFDMKFKYKMEFIY	179
Db	1680	YLYGIDSCVKNVYISFNLYT	1699	QY	181	DKQKHFDIYKQIEENPEFIDNIITKYLYSNEYSKDLBALNKYIEESINKITANGNDI	240
				Db	180	DRQKGFPIYKYSQINKPTVPTIDDIKSHLYSEVNRDETIVLESYRTNSLRKINSNHGDI	239
RESULT 3				QY	241	RULEXPADEDLVRLYNOELVEWNLAAASDILIRISMLKEDGGVYLDVLDILPGIQDLPKS	300
ID AAR95016				Db	240	RANSUFTTEOELLNTIYSQELLNENGLAAASDILVRLALKNFGGVYLDVMDLPIHSDLPKT	299
XX AC AAR95016;				QY	301	INKPSITNTSWEMIKLEAIMKYKEYIPIYTSKFNDFMLDEEVQSFESALSSEKSKSEIF	360
XX 16-OCT-2003 (revised)				Db	300	ISRPSSIGLDRWEMIKLEAIMKYKYINNYTSENFDKLDQQLKDNFKLIESKSEKSEIF	359
XX 08-JUL-1996 (first entry)				QY	361	LPDDIKVSPLEVKJAFANNSVINQALISLDSYCSDLVINQIKRYIKILNDLNPSE	420
XX C. difficile toxin A.				Db	360	SKLENLVNDSLEIKIAFALGVSINQALISKQSYLTNLVIEQVKRYQFLNCHLNPASIS	419
XX Toxin A; cytotoxin; enterotoxin; fusion protein; antitoxin; vaccine;				QY	421	GTDFNTMTKIFSDKLASINEDNMFMKITYLVKGFAPDVSRSTINLSGPGVYTGAYD	480
XX diarrhoea; diagnosis; therapy.				Db	420	DNMTDTTKIFHDSLFNSATAENSMFLTKIAPYLVQGFMPPEARSTISISGPGAYASAYD	479
XX Clostridium difficile; VPI strain 10463 (ATCC 10463).				QY	481	LLMFKDNSTNIHLPELENFPPKTKISOLTEQITSLWSFNQARAKSQFEEYKKGYPE	540
XX WO9612802-A1.				Db	480	FINLQENTIEKTLKASDLLEFPENNLSQLEQINSLSWSPDQASAKYQFEKYVRYDTG	539
XX 02-MAY-1996.				QY	541	GALGEDNDLDFACNTVLDKDY-VSKKILSS--MKTRNKEYIHYIVQLQGDKISYFASCNL	597
XX 23-OCT-1995; 95WO-US013737.				Db	540	GSLSEDNGVDENKNTALDQYLLNKKIPSNVVEEAGSKNYVHYIQLQGGDISYEATCNL	599
XX 24-OCT-1994; 94US-00329154.				QY	598	FSKDPYSSILYKQNEGSETAYYYVAD--ABIKIDKYRIPIYQISNKNIKLTIGCHK	655
XX 16-MAR-1995; 95US-00405496.				Db	600	FSKNPKNSIIQRNM--NESAKSYFLSDGDBEILELKYRIPELKNKEKVVTFIGHGK	657
XX 14-APR-1995; 95US-00422711.				QY	656	SBENTDTFRANLDVDSLSSEIEILNAXADISPKYIEINLGCNMFSYSISAEZYPGKL	715
XX 07-JUN-1995; 95US-00480604.				Db	658	DEFNTSEFARLSVDLSLSNEISSFLDTIKLDISPKNVEVNLGCNMFSYDFNVEEYPGKL	717
XX (OPHI-) OPHIDIAN PHARM INC.				QY	716	LLIKIDRVSELPMSIQDSITVSANQYEVRIINEEGKREILDHSGKWINKEBSIIKDISK	775
PI Williams JA, Padiye NV, Kink JA, Thalley BS, Stafford DC;				Db	718	LSIMDKITSTPVDNKNISITICANQYEVRIINSEGRKELLASHGXWINKEEALMSLSK	777
PI Firca JR;				QY	776	EYTSFNPKNKIIVSKYLHELSTLQELIRNNANSDDIDLEKKVMLTECEINVASNDRQ	835
XX WPI; 1996-230603/23.				Db	778	EYIFFDSIDNKLKAKSKNIPGLASISEDIKTLDDASVSPDTKFLNNLKLNISSIGY	837
XX N-PSDB; AAR29248.				QY	836	IVEGRIEEAKNLTSDSINIKNEFKLIESISLVDLKHQGLDDSHFISFEDISKTENG	895
XX Fusion proteins comprising non-toxin protein and part of toxin - useful				Db	838	IYVEKLEPVKNITHNSIDLDIDFENLLENVSDELVELKKNLDEKYLISFEDISKNGST	897
XX to form anti-toxins against Clostridium botulinum type A, and C.				QY	896	FRIRFINKETGNSIFETETEKEIFSEVATHISKEISNICKDTIFDNVNGKLVKKVNLDAHE	955
XX difficile type toxins, and to treat C. difficile intoxication, partic.				Db	898	YSVRFINKNGESVYVETKEIFSKYSEHITKEISTIKNSIITDNGNLLDNIQDHTSQ	957
XX diarrhoea.				QY	956	VNTLSNAFFIQSLIEYNTTKESLNSLVAMQVYAQLFSTGLNTITDASKVVELYSTAL	1015
XX Claim 63; Page 290-302; 434pp; English.				Db	958	VNTLNAAFFIQSLIDYSSNKVDLNTSVKQVLAQLFSTGLNTIYDLSIQLVNLISNAV	1017
XX Clostridium difficile VPI strain 10463 toxin A (AAR95016), the product of				QY	1016	DETIDLLPLTSEGLPIATIIDGVISGAIKELSETNDPILLOEIPAKIGIYAVNUTAS	1075
XX the toxin A gene (AAR29248), is a potent cytotoxin that plays a direct				Db	1018	NDTINVLPITTEGPIVSTILDGINLGAAIKELLDEHDPDLLKKELEAKVGLVAINNLSI	1077
XX role in damaging gastrointestinal tissues and is associated with				QY	1076	TAVTSALGTASOFSILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETGA	1135
XX diarrhoeic disease. It can be obd. by expression in transformed E. coli				Db	1078	AATVASIVGIGAEVITFLPIAGISAGIPSLVNNELIHDKATSVVFNHSESKYGP	1137
XX hosts of portions of DNA that together cover the entire toxin A gene.				QY	1136	FTLDDDKIIMPODDLVLSEIDFNNSITLKGCEIWEAEGSGHLLTDDIDHFFSPSSITY	1195
XX Toxin A, and portions of it (see also AAR95014-15 and AAR95017), pref.				Db	1138	LKTEDDKLVPIDDLVISIDFNNSIKLGTCTNLAWEAGSGHVTGNDIDHFFSPSSISS	1197
XX expressed as fusions to polyclonal affinity tags or maltose binding				QY	1196	RKPLWSIYDVLINIKKPKIDPSKDLAVLPNAPNRVFGYEMGWTPGFFSLDNDGTGKLDRI	1255
XX protein, are used to raise avian antibodies useful as antitoxins or				Db			
XX diagnostics, and in vaccine prodn. (Updated on 16-OCT-2003 to standardise				QY			
XX OS field)				Db			
XX Sequence 2710 AA;				QY			
Query Match				Db			
Best Local Similarity				QY			
Matches				Db			
Matches 871; Conservative 366; Mismatches 440; Indels 42; Gaps 15;				QY			
1 MNLVNAQLOKMYKFRQDEYVAILNALSEYNNMSESSVVEKYKLKIDNNITDNYL				Db			
1 MSLISKEELIKLAY-SIRPRENEYKITLNLDEYNKLTNNENKYLQKXKLNESIDVFM				QY			
61 NTKYKSGRNKALKKFEYLTMEVLELKNLSLTPVEKNLHFIWGGQINDTAINYNOWKD				Db			
60 NKYKTSRRLNLSNKLKLEKLVILKNSNTGFVEKNLHFVWIGGEVSDIALEYIKQWAD				QY			
121 VNSDYTVKVFYDSNAFLNTLTKTIVESATNTLESFRENLDNPEFDYKNFKYKMEIY				Db			

Db 1198 HIPSLSYSAIGIETENLDFSKKIMLPLNAPS RVFWWETGAVPGLRSLNDGTRILDSYR 1257
 Qy 1256 DHVEGQYRYAFIADALTKUKPRYEDVNRINLDGNTSRSPVIVITEQIRKNSYS 1315
 Db 1258 DLYPGKPYRFYAFY-DYAITLKPVEDTNIKLKDQTRNIMPITITITNEIRNKLSYS 1316
 Qy 1316 FYSGGGSYSLSLSPYNNIDNLAVENDTWIDVDNVVKNITTESDIEIQGELIENLSKL 1375
 Db 1317 FDGAGGTYSLLSSYPISITINUSKODLWIFNIDNEVREISBENGTIKKGKJIKOVLSKI 1376
 Qy 1376 NIENKIILNNHTINFGDINESNRFLSFLSILEDINIIEIDLVSYSKYILLGNCMK 1435
 Db 1377 DINKKLIIGNOITDFSDIDNDKDYIFLCELDKDLIIIEINLVAKSYLLSGDKY 1436
 Qy 1436 LIENSIOQKIDHIGFNGHQKYPYSYID-NETKYNQFIDYSKKEGLFTAEFNSII 1494
 Db 1437 LISNLNTIEKINTGLD---SKNIAVYTDENSKYFGAI-----SKTSQKSII 1483
 Qy 1495 RNIYMPSPNNL-----FYSSKDL--KDIRIINKGDKLIGNYFKD---DMKVSLS 1541
 Db 1484 H--YKDSKNILEFYNDSTLEFNSKOFIABEDINVMFKDDINTITGKYVDNNTDKSIDFS 1541
 Qy 1542 FTIEDTNIKLNGVYLDENGVAQILKPMNNAKSGALNTSLSMNFLESINIKNIFYNNLDP 1601
 Db 1542 ISLVSKQVKNVGYLNEYSVSSVYLDVFKNSDGHNTSNFNLFLDNISFWKLPGE--- 1598
 Qy 1602 NIEBILDTNFTIISGNSIGOFELICDXDKNIQIYFINFKIKETSITYLVGNRQNLIVEPS 1661
 Db 1599 NINFDVYKFLVCKTNLGYVEFICDNNKNIDIVFGEMKTSSSKSTIFSNGRNVVVEPI 1658
 Qy 1662 YHLDDSGNISSTVNFQKLYGIDRVVNVKVIAPNLYT 1700
 Db 1659 YN-PDTGEDISTLDFSEPLYGIDRYINKVLIAPDLYT 1696

RESULT 4

AAW68387
 ID AAW68387 standard; protein; 2710 AA.
 XX
 AC AAW68387;
 XX
 DT 07-DEC-1998 (first entry)
 XX
 DE Clostridium difficile toxin A.
 XX
 KW Antitoxin; vaccine; cytotoxin; toxin A; intoxication; immunogen;
 KW pseudomembranous enterocolitis.
 XX
 OS Clostridium difficile.
 XX
 PN WO3808540-A1.
 XX
 PD 05-MAR-1998.
 XX
 PF 28-AUG-1997; 97WO-US015394.
 XX
 PR 28-AUG-1996; 96US-00704159.
 XX
 PA (OPHI-) OPHIDIAN PHARM INC.
 XX
 PI Williams JA, Thalley BS;
 XX
 DR WPI; 1998-230234/20.
 DR N-PSDB; AAV30560.
 XX
 XX Host cell containing recombinant expression vector encoding Clostridium
 PT botulinum type B or E toxin - useful to treat humans and other animals at
 PT risk of intoxication with clostridial toxin.
 XX
 XX Example 15; Page 220-230; 428pp; English.
 PS This is the amino acid sequence of Clostridium difficile toxin A, deduced
 CC from the coding region (see AAV30560) of the toxin A gene. Toxin A is a

CC potent cytotoxin that plays a direct role in damaging gastrointestinal
 CC tissues. Severe cases of C. difficile intoxication result in
 CC pseudomembranous colitis. This would be prevented by neutralising the
 CC effects of toxin A in the gastrointestinal tract. Examples are provided
 CC of the production of recombinant C. difficile toxin A in host cells and
 CC of the in vivo neutralisation of toxin A by antibodies against
 CC recombinant toxin A polypeptides. The invention specifically relates to
 CC recombinant proteins derived from Clostridium botulinum toxins (see
 CC AAW68389-400) and their use as immunogens for the production of vaccines
 CC and antitoxins
 XX
 SQ Sequence 2710 AA;
 Query Match 51.5%; Score 4467; DB 2; Length 2710;
 Best Local Similarity 50.7%; Pred. No. 5e-225;
 Matches 871; Conservative 366; Mismatches 440; Indels 42; Gaps 15;
 Qy 1 MNLVNAQLOQWVYVKPRIQDEYVAIINALBEEYHNNESSVWEKYLKLDKNNLTNYL 60
 Db 1 MSLISKEELIKLAY-STRPRENEYKITLNDLENKLTNNNENKYLQKLNESIDVFM 59
 Qy 61 NTYKSGENKALKKFKYLTMTVELELNKNSLTPVEKNLHFWIGQINDTAININQWD 120
 Db 60 NKYKTSNRRLSNLKDKILAEVILIKNSNTSPVEKNLHFWIGGEVSDIALEYIKQWAD 119
 Qy 121 VNSDVTYKVFDSNAPLINTLKKTIIVESATNNTLESFRENLDNDFDYNKRYKMEIY 180
 Db 120 INAEYNIKLYDSEAFVLTNKAIVESSTTEALQLEEEIQNPFQDNMKFKYKMEFIY 179
 Qy 181 DKQHFIDYKSOEENPEFIIDNIITKYLSEYKOLEALNKYIEESLNKITANNNDI 240
 Db 180 DRQKFIINYKQKINQKPTVPTIDDIILKSHLSEYNRDETIVLESYRTNLRKINSVHIDI 239
 Qy 241 RNLEXFADBDLVRLYNQELVERWNLAAASDIIRISMLKXEDGGVYLDVILPGIQDLPKS 300
 Db 240 RANSLFTEQELLNYSQELLNAGNLAASDIVRLALAKNFGGVYLDVNLFGIHSDLFKT 299
 Qy 301 INKPSDITNTGEMIKLEAIKMYKEYIPGYTSKNFMDLDEEVQKSFESALSCKSKSEIF 360
 Db 300 ISRPSSIGLDREWMIKLEAIKMYKYINNYTSNFPDKLDQQLKDNFKLIEESKSEKSEIF 359
 Qy 361 LPEDDIKVSPLVKIAPANNVINOALISLKDSCVLDIVINQIKRYKILNDNLNPSINE 420
 Db 360 SKLENNVSLDEIKIAFALGVSINOALISKOGSYLTNLVIEQVKRYOFLNQHLPALAES 419
 Qy 421 GTDENTTKIFSDKLASISNEDNMFMKITNLYLKVGFAPDVRSTINISGPGVYTGAYD 480
 Db 420 DNNFTDTTKIFHDSLFNGSATAENSMFLTKIAPYLOVGFMPPEARSTISLSGQAYASAYD 479
 Qy 481 LLMFKDNSTNIHLLEPELRNFEFPKTKISQLTEQEITSLWGFNQARAKSQEYKKGYPF 540
 Db 480 FINQENTIEKTLKASDLIEFKFPENNLSQTEQINSLSWSPDQASAKYQEKYVDRDYG 539
 Qy 541 GALGEDDNLDPQNTVLDKQY-VSKILSS--MKTRNKEYIHYIVOLQDKITSYASCNL 597
 Db 540 GSILEDNGVDVFNKNTALDKVLLNKNKIPSNNVEEAGSKNYVHYIIQLQDDI-SYEATCNL 599
 Qy 598 FSKDPYSSILYQKNIEGSETAYYYVAD--AEIKIDKYRIPQIKNENIKLTFIGHCK 655
 Db 600 FSKNPKNSIIIQNNM--NESAKSYFLSDGESILELNKYRIPERLNKREKVVTFIGHCK 657
 Qy 656 SEFNTDTFANLDVDSLSSSEIETILNLAADISPKEYIENLLGCNMFYSISAEETYPGKL 715
 Db 658 DEFNTSEFARLSDVLSNEISSFLDTIILKDISPKNVEVNLGCNMFSDVFNVEETYPGKL 717
 Qy 716 LKIKDRVSELMFSPISQDSITVSANQYEVREINEGKREILDHSGKWINKEESIIKIDSK 775
 Db 718 LLSIMDKITSTLPDYNKNSITIGANQYEVREINSEGRKELLAHSGKWINKEEAIMSDLSK 777
 Qy 776 EYISFNPKENKIIIVKSKYLHELSLLOEIRNNANSDDILEKVMLETECEINVAINDRQ 835
 Db 778 EYIFFDSIDNKLAKSKNIPGLASISEDIKITLLDASVSPDTKFINLNKLNISSIGDY 837

Db 991 IITNAY---IKERDKL-----SGIIYNIENSIIISGHESFDIIRSNQHEMGDLSTVEQPK 1042
 Qy 784 ENKIVISKYHELSITLOBRNNANSSDLDLEKKWMLTECEINVAS-----NIDRQIVE 838
 Db 1043 KPEFYVXSE-LSSAKSIFDDIKNK-YITDPETKENVLYHQLDSDIKERIAFLDISHYAYP 1100
 Qy 839 GRIBKXNLTSDSINYIKNEFKLESI-----SDSLYDLKHONGL 878
 Db 1101 GSLEKQLS-----GYVFSNDINIABVLLASYGVSHYGVVPAPSDKLLLELRHTK 1156
 Qy 879 DDSHFISPEDISKTEGFRIRFINKETGNSI---FIETEKEIFSEYATHISKEISNIKOT 935
 Db 1157 SNSEWI--EKITP---YVYDILSDNVSNVLRPPLSEQKKILNDIKLEISKVS---EQ 1207
 Qy 936 IPDNVNGKLVKKNLDAHEVNTNSAFFQOSLEYNTTK-----ESLSNLSV----- 983
 Db 1208 YFMKLTQKSSVIGIKYSVDPRYNENLFLSLPINQNLTLFFMYRYPEMLYDIHIGIEN 1267
 Qy 984 -AMKVQVYAQLFSTGLNTITDASKVVELV-----STAL-DETIDL 1021
 Db 1268 KANREFIYKFSKSSNLNDFLINDERVNLLEGILIKKYKYLSEIHRTLTNTSTFADISPL 1327
 Qy 1022 LPTLSEGLPIITATII-----DGVSLGAIAK-----ELSTNDPLLR 1057
 Db 1328 LQTI---CPSITITIKTEYVYGHQLTNMTVASVVKYDPSNLGAINSDKSVSDYPAH 1384
 Qy 1058 QEIE-AKIGIMAVNLTAATAIVTSAL-----GIASGSILL 1093
 Db 1385 TIVEQAKYNLLSWNDFYNTHASIWDTTARQKSTNIEFHPQSLFLFDRDSKGLGLSLLY 1444
 Qy 1094 VPLAGISAGISLVNN-----ELIQDKATKVIDYFKHISLAETEGATLLDDKLIIMPQDD 1149
 Db 1445 LDTGGYGGYQKLRHNDITASTLYQTK-----YNDNLKLSNRDDFFLRKTQRIITWSNE 1498
 Qy 1150 LVLSEIDPNNSITLKGKEIWAEGSGHTLTDI--DHPFSSPSIYRKPWLSIYDVLN 1207
 Db 1499 L-----GNRLKNAQLEVLK-----DPLTEGILYQRISSLLIITEVHSLAQOISS 1548
 Qy 1208 IKK-----EKIDFKOLMVLNAPNAPRVGYEMGTWPGFRSLDNDGTKLDD 1252
 Db 1549 FWRVTDNPFHCHDSLAQALTFIKNI-----TSNR-----NFSSLYSGSI--- 1589
 Qy 1253 RIRHY-BQFYWRYFAIADALTKLAPRYEDNVRINLDG---NTRSFIVPVITTEQI 1308
 Db 1590 -VKIYFSESLNNWKYIKLPLVQTSLLRDYLTPEKLSGTSGSLNMGHLVPV----- 1642
 Qy 1309 RKNLSYF-YSGGYSYLSLSPYNNIDNLVENDTWIDVNVKNITIESDIOQ--- 1363
 Db 1643 -----SFTYDIGV-----INGNRISSEST---DVKNKIRSLKINGDILQHYIN 1682
 Qy 1364 -----KGELTENILSKLNTIEDNKI-----ILN---NHTINFY- 1392
 Db 1683 THYLSBQTKIDIVDFLQDNTIKVLESIDIKPSEIQOPLHSILSRQKEHVKNLLS 1742
 Qy 1393 GDINE-SNRF-----ISLTFSLIEDINI-----DIPI---DLVSKSYKILLSGNCKWL 1436
 Db 1743 GLLDFSNKLRQGLSLATNVLNVNFKESKINSDEVTETVTDQGLRYRVDITRIVGL 1802
 Qy 1437 I---ENSSDIQKIDHIGENGHEQKIPYSYDNETKNGFTDYSK---KEGLTAFESN 1490
 Db 1803 TPEKINSLSALEHMN-----IDAIMSVIGLVQYARMKQNDNISAIHDHAG 1849
 Qy 1491 E-SIRNIYMPDSNLFYSSKDLADIRIINKGVK-LLIGNYFKDDMKVLSFTIEDTN 1548
 Db 1850 AVSDIKNIIV---DKFLGILUTLNNRVYPPGVSGASLEGFTISSGLEVACAS----- 1897
 Qy 1549 TIKLAGV---YLDENGVAQILKFMNNAKSALNTSNLNFLESINIKINFYNNLD----- 1600
 Db 1898 -RMGCTAGRYL--SNVAKVIL-----PLLDI--GINIWLSYDSSLNHAHAKAT 1939
 Qy 1601 PNIETI---LDTNF-IISGSNIGQPE---LICKDKNIOPIYFNFKIKETSYTLVY-- 1650
 Db 1940 TQIEVISTAIDVSFSSINTALSIGAIVAPPLAIA-----IVPITI-PSHEVKNYAVYVQ 1993

Qy 1651 -GNRQNLIVSEPSYHLD-----SGNISSTVINFSQKLYG---IDRVNKKVII 1694
 Db 1994 INERHKWLEAKYLDNGSAXVLSINKATGIIIDSNQVGLNIIYLDRENPIL 2047

RESULT 7

ABU48941
 ID ABU48941 standard; protein; 4688 AA.
 XX
 AC ABU48941;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #34468..
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS Ureaplasma urealyticum.
 XX
 PN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR WPI; 2003-029926/02.
 DR N-P8DB; ACA52811.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 76865; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *X. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of

Db 787 N-----TLNQNEHKINTLNE-----QNEHKINTWKEEYEDQMNTLNEQNEDEKMSLKEBY 837
 QY 995 STGINTTIDAS--KVVELVSTALDEITDILPTLSEGLPIATIIDGVSLGAAIKELSETN 1052
 Db 838 ENKINQNSWNEIKIDVNVNEYIEEDVKLVITDEKKQDFDEINVAHKAHEKE-----892
 QY 1053 DPLLQBEIAKIGIMAVNLTAATAIIVTSALGIASGFSILLVPLAGISAGIPSLVNNELI 1112
 Db 893 -----QILLTEM-----EELK 903
 QY 1113 LQKATKVID-YPKHISLAEAGFTLLDDKIMPODDLVLSIEDFNNSITLKGKEIWR 1171
 Db 904 CQ-RDNKYSLVYKYLKISICMIINIECCDIENEDIRREEYINNKGK-LK-EVEE 961
 QY 1172 AB--GSGGHTLTDIDHFFSPSITVRKPLMSYD---VLNKKKIDFSKDLWLPNAP 1226
 Db 962 KEHRHSSFNILKSKEKFFN-----SIEDKSHLKKHEKDL-SKCKEI--EBK 1009
 QY 1227 NRFGYEWGTPGFRSLDNDGTGLDRIR-----DHVEGQFYWRYPAFIADALI 1275
 Db 1010 NK-----KIKELNDIKLQDEILVYKQSNQAVDHHKKSWI-----LJK 1050
 QY 1276 TKLPRYEDTNRLNDGNTRSPI-----VPVITTEQIR-KNLSYSFYGGSGYSLSLS 1328
 Db 1051 DKSKEIKDKENQINVEKNEKDKKKDDIRILNEELVKYKI-----LYNLKED 1101
 QY 1329 PYNNID-LNLVENDTWID-----VNVRKNITTESDETQK-----ELIENILS-----1373
 Db 1102 PLLQNDLLSKIDINSITNEGVCVKIEEHILDYDEINKRSNLFQLKNEICSLTTEV 1161
 QY 1374 KLNIEDNKILANHTTNFYG-----DINESREFILTSILIEDINIIE 1417
 Db 1162 MELNNKKNELTEBNKLNLDVQGGKKLKQVEKQKEIEKLNKQLTKONKQIDELNEEVE 1221
 QY 1418 -----IDLVSKS-----KYLLSGCMKLIENSDDIQ--KIDH-----1449
 Db 1222 KLANENIELITYSNDLANKFKDMKENLWKLDEDEDNIKQKSKIDMEKEIKYREDEK 1281
 QY 1450 -----IGFNGEHQK-----IPYSYIDNETKINGF-IDYSKK--1480
 Db 1282 RNLNEINLKKQNEEDMCIKYENMKIYGDICVXEEMSLTYKETSLLKYQIKVKYDEKCS 1341
 QY 1481 -----EGLPTAFSNE--SIIRNIYMPDSNLFYSSKDLKDIIRINKGVKLLI 1528
 Db 1342 QYDEIRPQYDEKCFQYDEINKYKALLNI---NITNWKVDSKVDNRNNEIISVDNKVEGI 1398
 QY 1529 GNYEK-----DDMKVLSFT--IEDTNTIKLGVYLDENGVAQILKEM---1569
 Db 1399 ANYLKQIFELNEEIRLKBINKISLLYSNELNEKNSYDINMKHIOEQ-----LLEFKT 1453
 QY 1570 --NNAKSALNTSNLMN-----FLESINI-KNIFYNNLDPN---IEFILD 1608
 Db 1454 NKENEKIINTSOYSDAYKKSDSKLGAQFVDDVNIYGNISNNIRTNKYEMFDF 1513
 QY 1609 TNPIISGNSIGQPELICDKDK--NIOPYFINFKIKET-----SYTLVYGNRQLYVE 1659
 Db 1514 TNIEKNGMHSKYHILLENKPCMKIIVENENIKSSNKIIGLYNVSRYVYGLREDLCKE 1573
 QY 1660 ---PSYHLDGSGNISSTVINSQKLYGIDRYNKKVI 1694
 Db 1574 EIVPS-----KIGNISKNENNNKN-NTCDGYDEKVTI 1606

RESULT 9
 ABM68924
 ID ABM68924 standard; protein; 3043 AA.
 XX AC
 XX ABM68924;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Photorhabdus luminescens protein sequence #2021.
 XX

KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; biopesticide; virulence factor; disease model; plague;
 XX whooping cough.
 OS Photorhabdus luminescens.
 XX WO200294867-A2.
 XX
 XX 28-NOV-2002.
 XX
 XX 07-FEB-2002; 2002WO-IB003040.
 XX
 XX 07-FEB-2001; 2001FR-00001659.
 XX
 XX (INSP) INST PASTEUR.
 XX (CNRS) CNRS CENT NAT RECH SCI.
 XX
 XX Duchaud E, Taourit S, Glaser P, Prangeul L, Kunst F, Danchin A;
 XX Buchrieser C;
 XX
 XX WPI; 2003-148459/14.
 XX
 XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
 XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.
 XX
 XX Claim 2; SEQ ID NO 2021; 1205pp; French.
 XX
 XX The invention relates to the isolation of genes and their encoded
 XX proteins from Photorhabdus luminescens. The isolated sequences are
 XX sources of probes and primers for detecting the genome of P. luminescens
 XX and related species; to study polymorphisms; for gene analysis and for
 XX detection/amplification of the genes. Antibodies (Ab) raised against the
 XX polypeptides encoded by the genes are used for detection/identification
 XX of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
 XX carry a gene-containing vector are used to select compounds that
 XX modulate, regulate, induce or inhibit expression of the genes in plants,
 XX animals or microorganisms other than P. luminescens and are able to alter
 XX response or sensitivity to toxins and antibiotics produced by P.
 XX luminescens. Cells transformed to express the genes are useful for
 XX recombinant production of the proteins, particularly toxins and
 XX antibacterials useful as insecticides, bactericides and fungicides. The
 XX genes, proteins, vectors containing the genes and Ab are also useful
 XX therapeutically (to treat microbial infection by bacteria or fungi that
 XX are sensitive to P. luminescens-encoded toxins or antibiotics) and as
 XX biopesticides. Other uses of the genes and the proteins are as virulence
 XX factors and for identifying targets of human diseases for which P.
 XX luminescens is a model (particularly plague and whooping cough). This
 XX sequence represents one of the isolated P. luminescens proteins
 XX
 XX Sequence 3043 AA;
 XX
 XX Query Match 4.3%; Score 374.5; DB 6; Length 3043;
 XX Best Local Similarity 18.5%; Pred. No. 1.8e-10;
 XX Matches 285; Conservative 262; Mismatches 578; Indels 419; Gaps 68;
 QY 81 MEVLELKNNSLTPVEKNLHFTWIGQINDTAINVQKNDVNSDYTVKVFYDSNAFLI-N 139
 Db 1022 IEYTKQNSISSSSVSAYFLGYNG-----LNQANVPAVVDPKHVAAGSFLFTG 1072
 QY 140 TLK--KTIVESATNNTLESFRENLDNDFYKFKEMEIIYDKQHFIDYKQSQBEEN 197
 Db 1073 TLSSGSLVVTSLDDNTYRVYHDG-----RFNSSLLYDNVVMVADYKDYQISDT 1120
 QY 198 P-----EFTIDNIKTYSNYSKDLALNKYIEESLNKTIAN-NGNDI--RNLEKF 246
 Db 1121 AGGFAAAVMQVNGEWQLVFQRYQYQYQYQYQYQYQYQYQYQYQYQYQYQYQYQYQY 1180
 QY 247 ADEDLVRLYNQELVERWNLAASDLIRLSMLKEGGVYLDVLDILPGIQPDLFKSINKEDS 306
 Db 1181 A-----TYREQVHQQLKTVATQ--FGVSVEGVTDGVYTEGEFSP-----1217
 QY 307 INTTSMWIKLEAIMKYKEYIPGYTSKNFDMDEEVQSFESALSCKSKSEIFLPLDDI 366

Db 1218 -----EHPATAAATRLREAAQKANKNDPVAUKLYQ-LQAPRTSSD----- 1259
 QY 367 KVSPLVKIAPANNVINQALISIKDSYCSDLVINQIKRYKILNDNLNPSINEGDTFT 426
 Db 1260 -----TALIDQIKQL-----NLQDYVYRAQYD---PILREASSVEK 1293
 QY 427 T-----MKIFSDKLASINEDNMFMKIKITNLYKVGAPDVRSTINLSGPGVYTGAVQDL 481
 Db 1294 TMLWQOIKAKGSAVAVQIDTALQGEERTSSVGERAAAEVYORGACGTG----- 1344
 QY 482 LMFKNSTNIHLLEPELNFPPKTKISQLTEBITSLWSPNQARAKSQEEYKGYFEG 541
 Db 1345 ---FNDG-----LNFR--EIKIPGVD-----KMSVLEMKRPLEG 1376
 QY 542 ALGEDNDLDFQNTVLDKDYVSKILSMKTRNKEYIHVIVQLQDKISYEASCNLFSD 601
 Db 1377 KLTHS-----OREALSARI---AETISQAEYI-----DKVLRQTA-MPSD 1412
 QY 602 PY--SSIL-----YQNIQEGSETAYYYVADAEIKEDKVRIPYQISNKNKILTFIG 652
 Db 1413 FHDAGSVNRLVPQDFYLSLVGDRGTGRCYPL-----VRTWAVALAG 1454
 QY 653 HGKSEFNTDTFANLDVDSLSSEITILNAKADISPKYIEINLLGCMFS--YGISAEET 710
 Db 1455 GGEAGINS-----LVQKLPASADPQAGSSTLLKNSLISLHYNVEAVQA 1498
 QY 711 YPKLLKIKIDRVSELMPSISQDSITVSANQVEVRINE---EGKRE-----ILD 756
 Db 1499 STELGQFNLSVSVRLAATGTSFALNTQNHMMVGTVAEGRYVYFDPNVGIFAFD 1558
 QY 757 HSGKWK--EBSIKDISKEYISFNPK-----NKILVSKYLHESTLLQETRNANS 810
 Db 1559 NT-KSLSRAMEQHLVVRGLAAHYGFGSKSPAFNLIEIDTRKMAEV-----PVGNGLV 1612
 QY 811 SDIDLEKVMTECEINVASND--RQI-----VEGRIEBAKNTSDSINYIKNEFKLIE 863
 Db 1613 AD-----LTFRE-ELASVIGQRQVEQVVSQVQWEMEDLRSLPALKTFDASQWARG 1663
 QY 864 STSDSLYLKHQGLDGHFI---SFEDISKTEGFRIRFINKE-----TGNISIF 910
 Db 1664 EVASARLAQEHQI---DSRWLPIATTD--QGEGRYAVQFINRDQPEOTRWLITSSTF 1718
 QY 911 IETEKEIFSEYAT--HISKEISNKTIFDNVNGKLVKKNVLDAAHEVNTLNSAFFIQS 967
 Db 1719 VBFRRFIDEHMAVLNEHFTLE-----HGRMLRGVGEAAPVDGLNAGFAVQT 1766
 QY 968 LIENYTKES-----LSNLSVAMKVQVYQAQLFSTGLNTITDASKVVELVSTAL----- 1015
 Db 1767 LIQWFADKNRHDAAAGSVISFDLATALKVHSLNVLQMAHGGVQDIATKVTALVTRLARGEV 1826
 QY 1016 -----DETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSETNDPILROEIEAKIGIM 1067
 Db 1827 VAAETSLDKDPASNLGHTVNEGA-----GVLFQGMVGLDAYE--LAHAENDVQKAVF 1876
 QY 1068 AVNLTAASTAVITSALGIASGF-----SIL---LVPLAGISAGIPSLVNNELILODKA 1117
 Db 1877 GCLAFDSASFYTGAGVAGLGVAGTAGAVLGASVILGLGAVFTALAQAFGAEDA 1936
 QY 1118 TKVIDYFKHISLAETEGATLLD-DKINPQDDVLSEIDFNNSNITGKCEIWRAGEGS 1176
 Db 1937 KAVGRYFDVDRKAYKNGVRYDNKNKVLPLAGAVIKSLDKRKHQISFDSQVIVRTHSGS 1996
 QY 1177 GHTLTDIDHFF---SSPSITY-RKPWLSIYDVNAIKEKIDFSKDL-----MVLNPA 1225
 Db 1997 --TGSGEINYYFWGDFPMVHDROQAIEVRSGIGYK-----DISPLHESDSNVILPGT 2050
 QY 1226 PNRVFGYMGWPGFRLSDNGTKLLDRDHYEGQVYRWYFAFIADALITLKPVEYEDT 1285
 Db 2051 PKSISYEWMLPLGATRHADGAFDVIRLLEE--DKRFDYDFVIFGGEETIRIHEHYDT 2108
 QY 1286 NVRLINDGNTRGFVPIVTEIQIRKNLSYFYGSGGYSLSLSPYNNMIDLNVENDT-- 1343

Db 2109 PIEVLDQRNLQVVPPELPE--LHSFLRYEIKGAGGEYLIGL-----NEGTQVKLTSDSASA 2164
 QY 1344 ---WVIDVDN-----VVKNTTIESDEIQGELIENILSKLNIEDKNKILNNH 1387
 Db 2165 LSRWIIDSSQIASDSISVKNQLVVGVVVELDPTQGVVL----- 2205
 QY 1388 TINFYCDINESNRFLISLTFISLEIDINIIEIDLVSYSKILLSGNCKMLIENSIDIOOKI 1447
 Db 2206 VVNGKGEVREVD-FAGLTQVQVSE-----DAGKQVPG--QQIEQHLSDLAK-- 2249
 QY 1448 DHIGNGEHQKYPISYIDNETKNG-----FDYSKKEGLFT 1485
 Db 2250 -----AHLHGQVYVVENYS-HNGRDVGRAFYDVGKERMLFT 2285

RESULT 10
 ABM66965
 ID ABM66965 standard; protein; 2410 AA.
 XX
 AC ABM66965;
 XT DT 20-NOV-2003 (first entry)
 XX Photorhabdus luminescens protein sequence #62.
 XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; biopesticide; virulence factor; disease model; plague;
 KW whooping cough.
 XX Photorhabdus luminescens.
 XX W0200294867-A2.
 PN 28-NOV-2002.
 PD
 XX 07-FEB-2002; 2002WO-1B03040.
 PF
 XX 07-FEB-2001; 2001FR-00001659.
 PR (INSP) INST PASTEUR.
 XX (CNRS) CNRS CENT NAT RECH SCI.
 PA
 PA
 XX
 PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
 PI Buchrieser C;
 PI
 XX WPI; 2003-148459/14.
 DE
 XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
 PS
 XX Claim 2; SEQ ID NO 62; 1205pp; French.
 CC The invention relates to the isolation of genes and their encoded
 CC proteins from Photorhabdus luminescens. The isolated sequences are
 CC sources of probes and primers for detecting the genome of P. luminescens
 CC and related species; to study polymorphisms; for gene analysis and for
 CC detection/amplification of the genes. Antibodies (Ab) raised against the
 CC polypeptides encoded by the genes are used for detection/identification
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
 CC carry a gene-containing vector are used to select compounds that
 CC modulate, regulate, induce or inhibit expression of the genes in plants,
 CC animals or microorganisms other than P. luminescens and are able to alter
 CC response or sensitivity to toxins and antibiotics produced by P.
 CC luminescens. Cells transformed to express the genes are useful for
 CC recombinant production of the proteins, particularly toxins and
 CC antibacterials useful as insecticides, bactericides and fungicides. The
 CC genes, proteins, vectors containing the genes and Ab are also useful
 CC therapeutically (to treat microbial infection by bacteria or fungi that
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
 CC biopesticides. Other uses of the genes and the proteins are as virulence
 CC factors and for identifying targets of human diseases for which P.
 CC luminescens is a model (particularly plague and whooping cough). This

Qy	225	I	E	S	L	N	K	I	T	A	N	G	N	D	I	R	N	L	E	K	F	A	D	E	D	L	V	R	Y	N	O	E	L	V	E	R	W	N	I	A	A	S	D	I	L	R	I	S	M	L	K	-----	278					
Db	288	I	E	K	T	L	T	H	L	T	O	T	R	G	V	D	H	-----	N	P	A	L	S	Q	A	R	D	S	E	L	K	T	L	A	P	M	-----	325																				
Qy	279	---	E	D	G	V	L	D	V	I	L	P	G	I	O	P	D	L	F	---	K	S	I	N	K	P	D	S	I	T	N	S	W	-----	E	M	I	K	A	I	M	K	Y	K	E	-----	325											
Db	326	S	G	S	E	P	R	O	V	S	A	N	I	H	O	F	---	A	A	D	P	V	I	F	A	K	S	T	I	N	T	E	T	L	R	S	G	F	P	A	E	G	N	A	C	I	V	K	I	E	S	G	L	E	Y	E	-----	384
Qy	326	Y	-----	I	P	G	Y	T	S	K	N	F	D	M	L	D	E	E	V	O	R	S	F	S	A	L	S	S	K	D	K	S	E	I	F	L	P	D	D	I	K	V	S	P	-----	370												
Db	385	Y	T	D	R	S	A	N	D	V	A	A	G	A	N	S	I	P	A	Y	-----	F	L	D	-----	Y	D	S	A	-----	D	Q	A	N	A	S	P	-----	418																			
Qy	371	L	E	V	K	I	A	P	A	N	N	S	V	I	N	O	A	L	I	S	K	D	S	C	D	V	I	N	O	I	K	R	Y	K	I	L	N	D	---	N	L	N	B	S	I	N	E	G	T	D	F	N	T	M	K	-----	429	
Db	419	A	Y	D	I	---	P	K	A	I	A	E	S	F	L	T	G	L	S	G	S	V	V	T	S	L	D	A	N	T	V	R	V	H	D	G	R	V	N	S	L	-----	465															
Qy	430	I	F	S	D	K	L	A	S	I	N	E	D	N	M	E	M	I	T	N	L	V	K	F	A	D	P	R	S	T	I	N	L	S	G	P	Y	V	T	G	A	Y	O	D	L	L	M	F	K	D	N	S	T	-----	488			
Db	466	L	Y	N	V	V	A	V	D	K	Y	---	V	A	A	T	T	S	L	A	V	A	Y	Q	Y	N	D	-----	E	W	Q	L	V	P	O	R	O	B	Y	Q	-----	511																
Qy	490	N	I	H	L	E	P	E	L	R	N	F	E	P	---	K	T	K	I	S	O	L	T	E	O	E	I	T	S	L	M	S	F	N	Q	-----	A	R	A	K	S	O	F	E	E	Y	K	Y	G	F	E	-----	540					
Db	512	D	G	O	R	L	R	L	K	G	D	E	E	P	L	S	I	Q	I	A	D	S	O	V	T	E	R	N	O	A	P	A	T	Y	R	E	Q	V	H	O	L	K	K	V	A	T	Q	G	V	S	E	G	I	A	D	-----	571	
Qy	541	G	A	L	G	E	---	D	D	N	L	D	P	A	Q	N	T	V	L	D	K	Y	V	S	K	I	L	S	M	K	---	T	R	N	K	E	Y	-----	I	H	Y	I	V	O	L	Q	-----	585										
Db	572	G	V	T	E	G	F	S	P	D	H	A	A	N	T	K	L	H	A	E	---	V	D	K	I	N	A	I	K	O	L	V	D	R	K	S	K	U	L	E	R	O	R	N	T	S	O	H	N	I	D	Q	Q	-----	630			
Qy	586	G	D	K	I	S	---	Y	E	A	S	C	N	-----	L	F	S	K	D	P	Y	S	I	L	L	Y	K	N	I	-----	E	G	S	-----	615																							
Db	631	I	K	I	N	I	T	L	E	Y	K	A	Q	N	T	V	L	R	E	A	G	F	E	O	S	L	W</																															

PS Claim 18; Page 779-781; 2188pp; English.

XX CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides (II), given in AAH81454 to AAH83120, from *Staphylococcus epidermidis*. (I) and (II) can have antibacterial activity and therefore can be used in vaccination. The nucleic acids (I) may be used to produce the S. epidermidis polypeptides (II) via the production of vectors containing them which are used to produce host cells which express the polypeptides. The polypeptides (II) (and/or nucleic acids) may then be used to vaccinate subjects and to raise antibodies against the bacteria. The polypeptides may also be used to assay for other inhibitors of their activity and therefore identify compounds that may be used for the treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to AAH55030 represent specifically claimed S. epidermidis genomic DNA polynucleotide sequences from the present invention. AAH55031 to AAH55098 represent oligonucleotide sequences and primers which are used in the exemplification of the present invention. N.B. The present invention specifically claims all the polynucleotide sequences given in the sequence listing of the present specification, however the sequence listing only goes up to SEQ ID NO:4454 so even though sequences are given in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present for SEQ ID NO:4455 to 4464

XX SQ Sequence 5024 AA;

Query Match 4.1%; Score 355.5; DB 4; Length 5024;
 Best Local Similarity 19.6%; Pred. No. 3.4e-09;
 Matches 369; Conservative 310; Mismatches 626; Indels 579; Gaps 101;

QY 20 QDEYVAILNALEEYHNSESVEKYLKLDINN-----LTDN-----YLN-- 61
 DB 2611 QMD-----TILNHI--FSAPTRSOVGEKIASQALNNTMKALRDSIADNELLQSSKYFNE 2665

QY 62 TVYKSGRNKALKFKEYITMEVLEKNNSLTPV-----EKNLFHFIWIG 105
 DB 2666 SEQQAYNQAVNAKNI-----NDQPTFVWANDIEQSVLNEVKTKDNLH---GD 2713

QY 106 Q--IND-----TAINYNQWKNDSYTVKVFYDYN-----AFLINTLKKT 144
 DB 2714 QKLANDKTDQAATNALNVLNQAQGNLETKVQ---NSNSRPEVQKVQVQLANQLNDAMKK 2770

QY 145 IVESATNNTLSEFRENLDNPEFDYKFKRMEIYDKQHFIDYKQSIQIE--NPEF-- 200
 DB 2771 LDALTAGN--DAIKQTSYINEDTSQ-----QVNFDE---YTRGRKNVAEOTNNPNSP 2819

QY 201 -----IIDNIKTYSNYSKDEALNKYIBESLNKITANNNDIRNLEKPADDLVRLY 255
 DB 2820 TWINTIADKI-----TEAKNDLHVQKL--EQAQQQSININQMTGLNQAKQEQ-- 2867

QY 256 NOELVERWNLAAASDILRAISMLKEDGGVYLDVILPG:QPLFKSINKPDSITNTSWEMI 315
 DB 2868 NGEIQOTQTRSEVHVQVINKAALNDS-----WNTLRQSITDEHEVKQTSNVIN 2915

QY 316 KL-----EAIMKYEIPYTSKNFMDLDEEVQSFESALSKSDSKSIFLPLDDIK 367
 DB 2916 ETVGQGTATNNAVDVKQIQTINQTSNPTWNL--EVERATSNVKT--DALGERELNDNK 2972

QY 368 VSPLEVKLAFANVS--INQALISLSDSYCSDL-----VINQIKRYKTLN----- 411
 DB 2973 NSK-----TFVYNHLNLDNQA--QKEALTHEIQATIVSQVNNIYNKAKALNNDMKKLK 3024

QY 412 -----DNLNPSINEGTQNTMTKIFSDKIASISNEDNMWFI-----KITNLYKVG 457
 DB 3025 DIVAQODNVQSNVYNEDSTQNNYND--TINHAQSIIDQVANTPMTSHDEIENAIN-- 3079

QY 458 PAPDVRSTIN--LSPGPGVYTGAYQDILLMPKDNSTNTHLEPELRNPEFP--KTKISQLTE-- 513
 DB 3080 ---NTHKAINALDGE---EKLQAKENA---NLLINSINDLNAPORDAINRLNVEA 3126

QY 514 -----QBITSLWSEFNOA-----RAKQF---EEYKKGYPFEGALGEDDNL 549
 DB 3127 QTRKVAEQLOQAALNDAMKHLRNSIQNSSVROESKYINASDAKQCYNHAVREVENI 3186

QY 550 DPAQNTVLDDKDYVSKKILSSMKTRKEYIHYIVLQGGDKISYEASC----- 595
 DB 3187 INEQPTLDKEII--KQLTDAVNQANN--LNGVELLDADQKQAHQSIPTLMLHNOAQNAL 3244

QY 596 -----NLFKSDPYSSILYQ-----KNIGSETAYYYYVADAB--IKEIDKIRIPYQIS 641
 DB 3245 NEKINNNAVTRAKVAALIGQAKILDHAMENLEES-----INDKEQVQSSNY-----IN 3292

QY 642 NKENIKLIF--IGHGKSEFNTDTANLDVDSLSSEIETILNLAKADISPK---YIEINL 695
 DB 3293 EDPDVQETVNNAVDHTVEILNQTVMNPTISIEDIHAINEV--NOAKQLRGKQKLYQTIDL 3351

QY 696 LGCNMFYSISAEETYPGKLLKIKDRVSELMPSISQDSITVSANQYEVRIEENKGR--- 752
 DB 3352 -----AKELSKLDDLSQSSSISNQIYIAKTTEVAQAIEAKAKSLNH 3395

QY 753 -----EILDHSGKWINK-----ESIIMDISKEYISFNPKENKIIVSKK 792
 DB 3396 AMKALNKIYKNAKVL--SSRPINEDQEKAYQAOAHNVDSIIHROTNPMPDPTVINS- 3453

QY 793 YLHELSTLLOETRN-----ANSSDDIDLEKXWMLTECEIN-----VASNI 832
 DB 3454 ITHELETAQNHLHGDKLAHAKQDAANVINGLIHLNVAQREVMINTNTNATREKVAKNL 3513

QY 833 DR-QIVEGRIEBAKNLSDSINYIKNEFKLIESISLSYDLKHQGLDSDSHFISPEDISK 891
 DB 3514 DNAQALDKAMETLQOVVAHK--NNILNDSKYL--EDSKYQQQYDRVIADABOLLNQTTNP 3570

QY 892 TENGFRIRPINKETGNSPIETEKEIFSEYATHISKEISNICKDTIFDNVNGKLVKKNLD 951
 DB 3571 TLEPYKVDIVKDNV-----LANEKILFG--AEKLSYDKSNANDEI----- 3608

QY 952 AAHEVNTLNSAFPIOSLIEYNTKESLNSLVAMKVQVYQLFST---GLNITDASKV 1008
 DB 3609 --KHMNYLANA-----QKSIKDMISHAALTEVKQLLOQAKTLDKAMKLEDKTVV 3659

QY 1009 ELVSPAL-----DETIDLLPTLSEGLPIIATII--DGVSLG--AAIKELSETNDP 1054
 DB 3660 -ITDTLTPNYTBASEDKKEKVD--QTVSHAQAIIDKINGSNVSLDQVQALFQASEN 3716

QY 1055 L---LRQETAEKI-----GIMAVNLTAASTA--IVTSALGIASGFSILLVPLAGISAG 1102
 DB 3717 LDGQDQVE--EAKVHANQTIQDLTHLNS--QQQTAKESVKNATKLE----- 3760

QY 1103 IFSLVNNELIQDKATKVIDYFKHSLAETGAFLLDDKIIIMPQDDLVLSIDFNNNSI 1162
 DB 3761 IATASNALALANKWKLQBFINHADSIENTSNYRQADDKIIAYDDALEHQDQIKSNA 3820

QY 1163 TLGKEIWEAEGSGHTLTDIDHF--FSSP--SITRKPKWLSIYDVNLKK-----EKIDPS 1216
 DB 3821 TQNEAKAQLQLINAEISLNGFERLNHAPRALEYIK---SLEKINNAQSALEDKVTOS 3877

QY 1217 KDLMLVLPNAPNRVFGYEMGWTGFRSLDNDGTGLLDRIDHVEGQFYWYFAFIADALIT 1276
 DB 3878 HDLLELEH-----LVNEGTLNLDIMGE-----LANAIVN 3906

QY 1277 KLKPRYEDTNVRINDGNTSRPIVPIVITEQIRK--NLSYFSGSGSGSYSLSPYNNMID 1335
 DB 3907 ----NVAPTKASIN-----YINADNRKONFTQAINNARDALN---KTQGNLD 3948

QY 1336 LNLVENDTVIDVD-----NVVKNTIDESIQKGLIENI--LSKMI--EDNKIILKN 1386
 DB 3949 FNAI--DTFKDDIFKTKDALNGIERLTAAKSKAEK--LIDSLKFINKAQFTHANDEIMNT 4004

QY 1387 HTINFVGDINESNRFLSIFLSILEDINIIEIDLVSKYSKILLSGNCKMLIENSSDIQOK 1446
 DB 4005 NS-----IAQLSRIVNQAFD--LNDAMKSLRDLNNAQFPVOASN---YINSDEDLKQ 4054

QY 1447 IDHIGFNGEHQKYPYPSYIDNETKNGFYDYKCKGLFTAEPFSNESIIRNIYMPDSNNLF 1506
 DB 4055 FDHA-----LSNARKV-----LAKENG----- 4071

QY 1507 IYSSKDLKDIRIINKGDVKLLIGNVFKDDMKVSLSFTEIDTNTIKLNGVYLDENGVA--- 1563

Db 4072 ----KNLDEIQI-----EGLKQVIEDTKD-ALNGI:QLRSKAKAKAI 4107
QY 1564 ---QILKPMNAKSAINTSN-----SLMFELESINIKNIPYNNLDNIFILDTNFI 1612
Db 4108 QYQSLSYINDAQRHIAESIHNSDDLSSLAITLSKA-----SOLDNAMKDLRDT--L 4158
QY 1613 ISGSNSIGQPELICDKDKNTQPYF 1636
Db 4159 ESNSTSVPNVSNVINADKNLQIEF 4182

RESULT 12

ABU42656
ID ABU42656 standard; protein; 6641 AA.

AC ABU42656;
XX 19-JUN-2003 (first entry)

DE Protein encoded by Prokaryotic essential gene #28183.
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX Staphylococcus epidermidis.

OS WO200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00915242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.
FA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Travick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

DR WPI; 2003-029926/02.
DR N-PSDB; ACA46526.
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 70580; 1766pp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 5213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6641 AA;

Query Match 4.0%; Score 347; DB 6; Length 6641;
Best Local Similarity 19.4%; Pred. No. 1.4e-08;
Matches 354; Conservative 318; Mismatches 662; Indels 494; Gaps 97;
QY 20 QDEYVAILNALAEYHNHNSSESVVEKYLKLDINN-----LTDN-----YLN-- 61
Db 3989 QKD---TILNHI--FSAPTRSQVGEKIASAKQLNNTWKALRDSIADNNEILQSSKYFED 4043
QY 62 TYKSGRNKALKKKEKYLTMVELELKNLSLPV-----EKNLHFIWIGG 105
Db 4044 SEQQNAYNQAVNAKNII-----NDQPTPVMANDEIQSVLNEVKQTKDNLH---GD 4091
QY 106 Q--IND-----TAINYNQWKDVNSDYTVKVFYDSN-----AFLINTLKKT 144
Db 4092 QKLANDKTDQAATNALNLYNQAGRNLETQVQ--NSNSRPEVKVQVQLANQINDAMKK 4148
QY 145 IVESATNNTLESFRENLDPEFDYNNKFKYRMEIYDKQKHFIIDYKQSIIE--NPEF-- 200
Db 4149 LDDALITGN--DAIKQTSNVINEDTSQ-----QVNFDE--YDRGKNIVAEQTPNNMSP 4197
QY 201 -----IIDNIKTLYLSNYSKDLKALNKY-----IESLNKITYANNNDIRNLEKFADELDV 252
Db 4198 TNINTIADKI-----TEAKNDLHGQVKLQQAQCSINTI-----NOMTGLNQAKQEL- 4245
QY 253 RLYNQELVERWNLAASDILRISMLKEDGVYLDVILPGIQDILFKSINKPDSITNTSW 312
Db 4246 ---NOEIQOTQTRSEVHQVINKQAALNDS-----MNTLRQSIITEHEVKQTSN 4290
QY 313 EMIKL-----EAIMKYKEVPGYTSKNTFMDLDEEVQSFESALSCKSKSEIFLPLD 364
Db 4291 YINETVGNQATYNNADVQKVIINQTSNPTWNPPL--EVERATSNVKISK-DALHGERELN 4347
QY 365 DIKVSPLVYKIAFANNSV--INQALISLKSYSVCSL-----VTNQKNRYKILN---- 411
Db 4348 DNKSK-----TFAVNHLDNLNQA--QKEALTHEIEQATVSVQVNNIYNKAKALNNDMK 4399
QY 412 -----DNLNPSINEGTFNTTKIFSDKLASISNEDNMFM-----KITNYL 454
Db 4400 KKLKDIVAQDNVRQSNVYNEDSTPQNMVND---TINHAQSIIDQVANTPTMSHDIEENAI 4456
QY 455 KYGFPADYRSTIN--LSGGVYVTGAYQDILLMPKDNSTNIHLLEPELRNFEFP-KTKISOLT 512
Db 4457 N-----NIKHAINALDGE-----HKLQQAQKENA---NLLINSUNDLNAPODAINRLV 4501
QY 513 E-----QSIISLWSFNQA-----RAKQSF---EYKKGVFEGALGED 546
Db 4502 NEAQTREKVAEQLQQAALNDAMKHLRNSIQNSSVRQESKYINADAKKEQYNHVAREV 4561
QY 547 DNLDFAQNTVLDKDYVSKKILSSMKTRNKKEYIHYIVQLQGDKISYEASCN---LFSKDPYS 604
Db 4562 ENINEQHPTLQKEII--KQLTDGVNAQAND--LNGVELLDADKQNAHQSIPTLMHLNQAOQ 4619
QY 605 SILYQK-NIEGETAYYYVADAETKEIDKTRIPQISNKNRIKLTFFGHGKSEFWDTTF 663
Db 4620 NALNEKINNNAVTRTEVAAILIGQAKLHDAMENLEESIKDKEQVK-----QSSNYINEDSD 4674
QY 664 ANLVDVDSLSEIETILN-----LAKADISPKVIEIN-----LLGCNMFYSISAEETYP 712
Db 4675 VOETYNDAVDHVTIELNQTNTPTLSIEDIEHAINVQNAKKQLRGKQKLYQTIDLAD---- 4731

Qy	713	GKULLKIKORVSELMPSISODSITVSGANQVEVRAINBEGKR-----EILD 756
Db	4732	-KELSKLDLITSOQSSISINQIYTAKTRETEVAQAI EKASLHAMKALNKVYNADKVL D 4790
Qy	757	HSCKWINKE-----ESI IKDSSKEYISFNPKENKIIVKSYLHLSHTLLQERNN-- 807
Db	4791	-SSRFINEOPEKKAQOALNHVDSIIHRQTNPEMDPTVINS-ITHLETTAQNLLHGDOK 4848
Qy	808	-----ANSSDIDLEKKVMLTECEIN-----VASNIDR-QIVEGRIEEAKNJT 848
Db	4849	LAHAQQAANVINGLIHLNVAQEVMTINTWTATTREKVAKLNDAQALDKAMETLQQVV 4908
Qy	849	SDSINYIKBEFKLIESDSLYDLKHQNGLDDSHSFISFEDISKTENGFRIRPINKETGNS 908
Db	4909	AHK--NNILNDSKYL--EDSKYQQQYDRVIADAEQLLNQNTNPTLSEPKYVDIVKQNV-- 4962
Qy	909	IFETEKEIPSEYATHISKEISIMKOTIPDNVNGKLVKVNLDAAEVNTLASAFIQSL 968
Db	4963	--LANEKILFG--AEKLSYDKSNADEI-----KHMVYLANA----- 4995
Qy	969	IEYNTTKESLNSIVAMKVQ--VYAQLFSTGLNTITDASKVVVELVSTAL----- 1015
Db	4996	-QKQSIKMDISHAALRETEVKQLLQQA KILDEAMKSLDKTQVV-ITDTLPNTYEASEDK 5053
Qy	1016	DEITDILLPTLSEGLPIIATII-DGVSLG---AAIKELSEINDPL---LRQETEAKI--- 1064
Db	5054	KEKVD--QTVSHAQAIIDKINGSNVSLDQVRQALEGLTQASENLGDGQORVE-EAKVHANQ 5110
Qy	1065	---GIMAVNLTAASTA--IVTSALGIASGFSILLVPLAGISAGIPSLVNNELILQDKAK 1119
Db	5111	TIDQLTHLNSLQOQTAKESVKNATKLEE-----IAIVSNNAQALNKVMGK 5155
Qy	1120	VIDYFKHISLAETEGAPTLDDKLIIMPQDLVISEIDFNNSITLKGCEIWRABGSGHT 1179
Db	5156	LEQFINHADSVENSNDYRQADDDKIIAYDEALEHGDQIOKTNATONETK----- 5204
Qy	1180	LTDDIDHFFSPSITYRKPLMSIYDVNL-1KKEKIDPSKDLMLPNAPNEVFEGVMGWP 1238
Db	5205	-----QALQOIIVAEITSUNGFERLUNHARPALEYIKLEKINNAKSALEDKVTQSH 5256
Qy	1239	GPRSLD---NDGTKLDRIDRHDYEQGYWRYFAFIADALITKLKPRYEDTNVIRLNGNT 1295
Db	5257	DLLELEHIVNEGTLNDIMGZ-----LANAIVN---NYAPT KASIN----- 5294
Qy	1296	RSFIVPVITTEQIRK-NLSVSYFGSGGYSLSUSPYMNQIDMLNVENDTWDVIDV----- 1349
Db	5295	-----YINADNLKDRFTQAINNARDALN--KTQGNLDFNAI--DTFKDDIFKTKDA 5343
Qy	1350	NVKNKITESDEIQKGELIENI--LQKLNLI-ENKILNHNHTINFGYDINESNRRFISLT 1405
Db	5344	LNGLERITAAKSAEK--LIDSLKFINKAQFTHANDEIINTNS-----IAQLSRIVNQA 5395
Qy	1406	FSILEDINIIEIDLVSKSYKILLSGNCKMLIENSDDIOQKIDHIGFENGSHQKYPYSYI 1465
Db	5396	FD-LNDAMKSLRDELNNQAPFQVQASN---VYNSEDEKLQOFHALSNA--RKVL----- 5444
Qy	1466	DNEPKYNG-FIDYKSKGSLFTABESNESIERNIYMPDSNNLFIYSSKD-LKDIRINKGD 1523
Db	5445	---AKENGKNDLDEKIQIGL-----KQVIEDTKDALINGQRUSKAK 5481
Qy	1524	VKLLIGNYFKDMMKVSFTTIEDTNTIKLNGVYLDENGVAQILKFMNNAKSALNTSLSM 1583
Db	5482	AKAI--QYVQ-----SLSYINDAQRHAEANNIH-NSDDLSSLANTLSKASDLONAMKDLR 5533
Qy	1584	NFLSISINI---KNIPYNNLDPNIEFILD 1608
Db	5534	DTISNSTSVPSNVNINADKNLQIIEFD 5561

AC	ABP30314;
XX	
XX	24-JUL-2002 (first entry)
XX	
DE	Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3159.
XX	
KW	Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX	antibacterial; gene therapy.
XX	
OS	Staphylococcus epidermidis.
XX	
PN	US6380370-B1.
XX	
PD	30-APR-2002.
XX	
PF	13-AUG-1998; 98US-00134001.
XX	
PR	14-AUG-1997; 97US-0055779P.
XX	
PR	08-NOV-1997; 97US-0064964P.
XX	
PA	(GENO-) GENOME THERAPEUTICS CORP.
XX	
PI	Doucette-Stamm LA, Bush D;
XX	
DR	WPI; 2002-381255/41.
DR	N-PSDE; ABN90859.
XX	
PT	Novel isolated nucleic acid encoding a Staphylococcus epidermis
PT	polypeptide, useful for diagnosing and treating bacterial infections.
XX	
PS	Disclosure; SEQ ID NO 3159; 267pp; English.
XX	
CC	ABN905338 to ABN93374 represent Staphylococcus epidermidis open reading
CC	frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC	given in ABN35124 to ABP37960. The S. epidermidis sequences have
CC	antibacterial activity and can be used in gene therapy. The sequences can
CC	also be used in the diagnosis and treatment of bacterial infections,
CC	particularly S. epidermidis infections. The sequences can be used to
CC	screen for compounds able to interfere with the S. epidermidis life cycle
CC	or inhibit S. epidermidis infection. N.B. The sequence data for this
CC	parent did not form part of the printed specification, but was obtained
CC	in electronic format directly from the USPTO web site
XX	
SQ	Sequence 10182 AA;

CC of infection with P. falciparum. Furthermore, (i) (especially when they
 CC are rifins or secreted or membrane proteins) can aid the identification
 CC of drugs to treat or prevent P. falciparum infection, or they can be used
 CC to identify drug resistance in P. falciparum. Sequencing of the
 CC Plasmodium chromosome 2 and the subsequent identification of proteins
 CC encoded by it will help to expand our understanding of parasite biology,
 CC a process hampered by the complexity of the parasitic lifecycle, and
 CC provide new targets for vaccine and drug development. Parasite resistance
 CC to drugs and mosquito resistance to insecticides have led to a resurgence
 CC of malaria in many parts of the world, and there is a pressing need for
 CC vaccines and new drugs. AAA70078 to AAA70287 and AAA8144 to AAA81852
 CC represent nucleotide and protein sequences given in the present
 CC invention, but which are not specifically mentioned within the
 CC specification
 XX
 SQ Sequence 2485 AA;

Query Match 4.0%; Score 343; DB 3; Length 2485;
 Best Local Similarity 19.7%; Pred. No. 6.1e-09;
 Matches 356; Conservative 269; Mismatches 643; Indels 542; Gaps 91;
 QY 35 HNMSSSVVYKYLKLDINLFDN-----YLVTKYKGRKALKKF---KEYITMEVLE 85
 Db 157 HTNSNI---YINCNDNDIRNSNVQHYNDKIKENKQNKYVLINDYINNKYI- 212
 QY 86 LKNSLTPVEKHLFIWIGQINDTAINVQ-----WKVNSDYTVKYVDNAFLN 139
 Db 213 LSKNKTCKINKGKL-----IKKKVANSRRNHILYKCRNLKYNVFSDD- 260
 QY 140 TLKKTIVGATN--NTLESFRENAN--DPEPDYKFKYKRMELIYDKQHFIDYKSOI 194
 Db 261 -----IKSEVNVCSLTVLHKYINNMNDYLDN-----IHTNNSNIYD-----I 301
 QY 195 EENPEPIIDNIKT-----LSNEYSKDLALN--KYIEESLNKLTAN----- 235
 Db 302 NYTENVINSTCRYPIGNNTLSXDEVTXSSSKINSLSYDPIIN-VKNNDIPILHDK 360
 QY 236 NGNDIRNLEKF-----ADELVRLYNOELVER-----WNLAASDIIRISMLKEDGVY 284
 Db 361 NIIAISNESCHKEEKEKYMYNSLVVEEKQKKMIWN--SLNVLPIDLLKNG--- 414
 QY 285 LDVDILPGQPDLFSINKPDSITNTSMEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVR 344
 Db 415 -----HDEINKEICKKKKSPSQNDIKSKLYNKS-----SKS----- 450
 QY 345 SPESALSSKSDSEIFPLDDIKVPLEVKIAPANNVINOALISLKDSYCSDLVI----- 400
 Db 451 --EKVLYTNKNSNTPFIPIFLN-----KYGDKFKSENILYDMYNNKKVYIHDKLIYTM 504
 QY 401 --NQIKORY-----KILNDNLNPSINEGTFDNTM--KIPSD-KLASISNEDNMWMI 448
 Db 505 YSNKLAQKHVYSTNINLNNIGKYLNDGLHLSNNMYCRLNSNPPYKSISLNNNVFFY 564
 QY 449 KITNYLKVGFAPDRSTINLSPGVVTGAYQDLLMKD-NSNIHLEPE---LNPFP 504
 Db 565 K-----KRKSNNNNNNNISSSSSSKQNHVIINKKISSYNIHYKPKDSFENFLFF 620
 QY 505 KTKISOLTQETISLWSFNQAKAQOFERYKGYEGALGEDDNLDFQNTVLDKDYSK 564
 Db 621 KEKI--LPSKDTCTV-----FNRQKDLFE----- 643
 QY 565 KILSSMKTENKEYIHVIVQLQ--GDKISYEASCNLPSKDPY-----SSILY---QKNI 614
 Db 644 -----KSENIHKVCSFNNTSDISHSVNV--KKEPFPALKNSIRHIPPKNENIY 693
 QY 615 SETAYYYYVADAEIKDIKYPYQISNKNKTLFPIGHGKSEFNTDFANLDVDSLSSE 674
 Db 694 TSGKSNHYQDKETVLLKKK--KEINDKTFSSCLINHNIITYTLQNGVKN----- 744
 QY 675 IETILNLAQADISPKYIEINLCCNMFYSIS-----AEETYPKLLIKDQVSELMPS 729
 Db 745 -----LNMLGIRDSIYKIDKKNMLKECYNGNNDNNKKNKKKKKKKXLS 786

QY 730 ISQDSITVSANQYEVRIINEEGREILDHS-GKWINKESIIKDISKE-----YISFNP 782
 Db 787 FSCILINDNITPVESDEKKNNSNIKMSDIFNVVKKSNLYNNLSNRDSTVDMENKYS 846
 QY 783 KENKIIVKYKYLHELSTLQETIRNNANSSDIDLEKKWMLTECEI--NVASNIDRQIVEGR 840
 Db 847 EYINQRTNKIVELSN--KRIRNYKLYS-MDEIFKVLKEKKYIDNISNNMER----- 897
 QY 841 IEBAKNLTSDSINYIKNEF--KLIESISDLYDLKHQGLDSDSHFISPE-DISKTEGPR 897
 Db 898 -----VYI-KNEMINEKISKMDILLPCDKNKLNSMPCVPIENNISREN----- 942
 QY 898 IREINKETGNSPIETEK-----EIPSEYATHISEISNIKDTIFDNVNGK-----LVKK 947
 Db 943 -----EKNSSVILNKKKNENMFCVGRHLCHWGK--YNNQDNIVDQGNIKNEEIEITKH 994
 QY 948 VNLDAAEHVNTNSA-----FFIQSLIEYNTTRESLSNLSVAMKVQVIAQLFSTGLT 1000
 Db 995 DEYISREKKNYKSKIRNFDDYKYEQVLSYHTLDEDKK----- 1034
 QY 1001 ITDASKVVELSTALDETTIDLLPTLSEGLPIITATII---DGVSLGAATKELSETNDPLL 1056
 Db 1035 -NDMNLLIDMNEAIIETYN-----GVNNIILDRKNNNRKDKMEKEMEKEMKKM 1084
 QY 1057 ROIEAKIGIMAVNLTAASTAIVTSALGIASGFSILLVPLAGISAGIPSLVNNELIQOK 1116
 Db 1085 EKEMEK--VMEKEMEKVMEKEVEKEL-----KNEMNKNMKNMKNMKNMKN 1129
 QY 1117 ATK-----VIDYFKHISLAETEGFTLLDDKIIMPQDDLVLSEIDFN--NNSITLGCBEIWR 1171
 Db 1130 IYKXNEIYVDNDKELEIVNEE-----KKLIVFPN--YESDVHKMMMSININCK--- 1177
 QY 1172 AEGSGHTTDDIDHFPSSPTYRKFWLSIYDVNLKKEKIDFKDLMLVLPNAPNRVVG 1231
 Db 1178 -----DDVN-----NILKEYVDNS-----CLAQKEENI-- 1200
 QY 1232 YEMGWTPGSRSLDNDCTKLLDRIRDHYEQFYVRYFAF-----IADALIT 1276
 Db 1201 -----FRLPFLNKK--DKV-----WKRFNIKNKNTIHIHEEMKRIYOTINK 1241
 QY 1277 KLPK-----RYED--TNVRINLDNGTRSFIVPITTEQIRKNLSYFSGSGSYLSLS 1328
 Db 1242 NVFPIYNFREYENFLNHLTYNFPKNDLPKLSYKVSMMNIR-NLYTANKHINNIDYMNK 1300
 QY 1329 PYNNMI-----DLNLVENDTWI-----DVDNVVKNITIE----- 1358
 Db 1301 LYQNIVTLKYQVANIDNDHHICKGGGLDYINMNLISKECKNRKDKTYLKNIFHYKKKD 1360
 QY 1359 -----SDEIQKGLIENILSKLNIENDKIILANH-TINFYD-----INESNRFISL 1404
 Db 1361 ARFFINDEIGSDNYMDYDIKKYSNDENNYKLANEKOMNISNDEDMIPTLNSHGNNFPSC 1420
 QY 1405 TFSILEDINIIEIDLVSYSKILLSGNCKLIENSDDIQKIDHIGFNGEHQKYPYGY 1464
 Db 1421 QPNLEKKSYTIDNLNLYDSNMDDFTEBKYNFVNNDL-----FNTKWK---FNP 1469
 QY 1465 IDNETKYG-FIDYSKKEGLTAEFSESNEISIRIYMPDSNNLIYFYSKDKOIRINKGD 1523
 Db 1470 SKGKLNENKFFVNSNEDGVFSF-FKNMNLREL-NKSNLSKLESVKNSNNNSNKG 1527
 QY 1524 VKLLIGNY-----FKDDMKVSL-----SFTIEDTNTI--KNG---VYLDENG 1562
 Db 1528 DN--IGNMNNNTNTVTIASDEHISTKGDTHDSFSRDNDCILLKIEGRSKKYS- 1581
 QY 1563 AQILKFNNAKSAINTSNLSMNFLESINIKNFPYNNLDNIEFILTNTIISGS-NSIQ 1621
 Db 1582 ---ITLYNEDKSLNENDNETINEYV-----CSNIDVN-----EWEDKVGTCNSVD 1627
 QY 1622 FELICDKKN 1631
 Db 1628 KETEKANKN 1637

RESULT 15
ABU48947
ID ABU48947 standard; protein; 5005 AA.
XX AC ABU48947;
XX DT 19-JUN-2003 (first entry)
XX DE Protein encoded by Prokaryotic essential gene #34474.
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Ureaplasma urealyticum.
XX PN W0200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002W0-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 23-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haeelbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX DR WPI; 2003-029926/02.
XX DR N-PSDB; ACAS2817.
XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 25; SEQ ID NO 76871; 1766pp; English.
XX CC The invention relates to an isolated nucleic acid comprising any one of
XX CC the 6213 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX CC polypeptide or its fragment whose expression is inhibited by the
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX CC proliferation or the activity of a gene in an operon required for
XX CC proliferation; (7) identifying a compound that influences the activity of
XX CC the gene product or that has an activity against a biological pathway
XX CC required for proliferation, or that inhibits cellular proliferation; (8)
XX CC identifying a gene required for cellular proliferation or the biological
XX CC pathway in which a proliferation-required gene or its gene product lies
XX CC or a gene on which the test compound that inhibits proliferation of an
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX CC product is overexpressed or underexpressed; (12) determining the extent
XX CC to which each of the strains is present in a culture or collection of
XX CC strains; or (13) identifying the target of a compound that inhibits the
XX CC proliferation of an organism. The antisense nucleic acids are useful for
XX CC identifying proteins or screening for homologous nucleic acids required
XX CC for cellular proliferation to isolate candidate molecules for rational
XX CC drug discovery programs, or for screening homologous nucleic acids
XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
XX CC the target prokaryotic essential genes. Note: The sequence data for this
XX CC patent did not form part of the printed specification, but was obtained
XX CC in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pt_sequences
XX CC Sequence 5005 AA;
XX SQ

Query Match 3.9%; Score 340.5; DB 6; Length 5005;
Best Local Similarity 19.7%; Pred. No. 2.1e-08;
Matches 405; Conservative 283; Mismatches 676; Indels 697; Gaps 105;
QY 28 INALBEEYHNMSESSVVEKYLKLDINMLTD-----NYLNTYKSGSRNK-ALKK-- 74
DB 744 LNLMEY-----YLKLYIDNNGDVVSDPSINNANVEYFNLDPKALKSNR 792
QY 75 -----RKEYLTMEV-----LELKN-----NSLTPEKNLH 99
DB 793 IYKFDGLYFQVNTVTKITSANKVNMNITPQIQATSKINLNSPKVTINNITPASADLY 852
QY 100 F-----IWGGQINDTAINYNQWKDVNSDYTVK-VFYDSNAFLINTLKTIVESATN 151
DB 853 LPLVSTDDIPAKQDVIVTIG-----NKDNTKNDLMFTSNLEYDDIYKTVGWKIHAN 904
QY 152 NTLSEFRENLDPEFDYN---KFYRKM-----EIIYDKQK----- 184
DB 905 NL-----SPEVNYQIKSVKFRQKPVLGAFNVNNSDDNHVLIDSKQTPGSTPKASF 955
QY 185 -----HFIDYKSOIENPEFIIDN-----IKTYLSN 212
DB 956 DLVSVTASDVVDANVKTNSVHIVINDGTSLNKCAKIVYNDGHEHLVSNFVVLNIGVN 1015
QY 213 EYKXDLKALNKYIEESLNKI--TANNGNDIR---NLEKPADEDLVRLYNQELVERWNL 266
DB 1016 EYDFNLDLKHNRYSFKGLIYTEPNQNDYKFSPLNNQKITPTKTASVNEQLTIDTKLV 1075
QY 267 AASDI-----LRISMLKEDGGVYLDVD-----ILPGIQDLPKFSINKPD 305
DB 1076 RKPDSNLKINLQINDPNDFLEQDSILEITFHDEKDKTHNVIGKINDANNKKTLEF 1135
QY 306 SITNTSWMEMIKLEAIMKYKEYIPGYTSKNFDMLEDEQVRSFESALSKSKSEIFLPLDD 365
DB 1136 SVENTN--TFKIQNHKIVDININATKN-----KIQA--NAISNNSKNISY----D 1181
QY 366 IKVSPLEV-----KIAPANSVINQALISLKDYSVDIVNQIKRYKILNDNLNPSNEG 421
DB 1182 ASSNPSKILSTFNLFNVINVPNTNLPNTASIDV---ELKSSQNLKQYLRALYID 1238
QY 422 TDFNTTKIFSDKLASINED---NMFMKINYLKVGFPADVRSINLSGPGVVTGA 477
DB 1239 ---NNHQKIMSD-VASVNVVDLAHINLNFN-HLTPNRKYNPA-----GIY--- 1278
QY 478 YQDLMLFMDNSTNIHLEPELNFEPFKTKISQTEQITSLWSFNQARAKSQEYKKG 537
DB 1279 -----YFNDQON-----NDE---TQGGKIFVKEHSKTYDFSTLSTQLLEPKTH 1321
QY 538 YFEGALGEDD-NLDPANQNT---VLDKDYVSKILSSMKTRNKYIHIIVQL-QGDKISYE 592
DB 1322 NVD-----EDQFNVDNINDDQVLEPGMLVCLPENLVDKQPVLTQTTLVKKDEHSFS 1377
QY 593 AS---CNLFSKDPY---SSILYOK-----NIESETAYYYY----- 622
DB 1378 ASGVITNLADHTVRLTSVLSLKQPOLANVINNNNNNEILLNIEDNQKTIHTLSKIV 1437
QY 623 -----VADAEIKEDK-YRIPYQIS-----NKRNIKLTFFIGHGKSEFNTDFANLDV 668
DB 1438 NEISNLADAYPANNNGYDVKFNINIKNNKLNKVKVVFEDNEHQLISTN----- 1490
QY 669 DSLSEIEITLNLAKADISPKYIEINLLGCNMFYSISAEETYPKGL--LLIKK----- 720
DB 1491 DLLVNLKD-----QTNLSLQNFSS-----NLKPNHLVRLKLVYVGBQN 1531
QY 721 -DRUSE-----LMPSTSQDS-----ITVSANQYEVRIINEBGRILD-----HSGK 760
DB 1532 FDAINEQKNILANPSLNVSSFSFTPAKIKVSKNAIDVWQNALIKLILDDSDNQLHSGD 1591
QY 761 WINKESIIKDISKEYISFNPKENKIVKSKYLHELSTLLOEIRNNANSDDILEKKVM 820
DB 1592 EIN-----INRYIKGTQNIISTP-----ATISDQKR----- 1618

Search completed: April 1, 2004, 16:41:59

Job time : 95.25 secs

QY	821	LTEC-EINAGNIDROIVEGRIBAK-----NLTSOSINY--IKNEFKLIESIDS	868
Db	1619	YTKMATNLIALGDIYEIVSVIKQKYNVSPILFELPSGPIGYTLAPVFKTISLMDP	1678
QY	869	LYDLKHQGLDSDHFIISPEDISKTENGPRIPIFINKETGNSIFIEKEIFESEYATHISKE	928
Db	1679	VY--HEN-----KNTADLKINLRIENIGASLFNKDKIFIKR--KNDGKQ	1719
QY	929	ISNIKDTIFDNVNGKLVKKNLDAAHAEVNTLNSAFFIQSLIEYNTTKESINLSVAMKVQ	988
Db	1720	ISFI-----HKVTSANDANY-----EWE-FKDLLNREYTLERY	1752
QY	989	VYAQLFSTGLNTITDASKVVELVSTALDETIDLLPTLSEGLPIATIIDGVSLGAAIKEL	1048
Db	1753	VYKKNQFNQNSVESDVIDIWIYENLNKTKFLPTKPLGI-----IGAPIKEI	1801
QY	1049	SETN-----DPLLQREIEBAKIGIMAVNLTAATAIVTSALG---IASGFSILL	1093
Db	1802	SDNGAKVQLKPAINDFDDVLKENOTFKENIQPENENNANLNEISEHEGKVEIIDGQKFFV	1861
QY	1094	VPLAGISAGISLVNNEILQDKATKVIDYFKHISLAETEGAFLLDDKIIMPODDLVLIS	1153
Db	1862	ANLANNIK-----VNKEY--KVYKI--YFDE-NQDVGVGVY-----	1891
QY	1154	EIDFNN-----NSITLGKEIWRABGGSGGHTLTDIDHFPSSPSITYRKPLWSIYDVUNI	1208
Db	1892	KINFKNDYKEPNVYVY-----DASVNTTQTYVFTNKFAIASFSNN---LTDVDVAN-	1939
QY	1209	KKEKIDFGKDLMLVPAENRVPGY-----EMGWTGFRSLDNDGT*KLDRDRDH	1257
Db	1940	KONISINLDSRV-----ETIQGHFKAKYISNDRVVTWTIIPAPNIGNN-----KNN	1988
QY	1258	YEGQYWRYPAFIADALITKLPRYEDTNVRINLDGNTSRFIVPITTEQIRKNLSYSFY	1317
Db	1989	IALNFELNQNQLISNRLVTPAALYYSKD--VNVDEHANVWV-----IKNNVN---	2034
QY	1318	GGGSYSLSLSPYNNIDLN-----LVENDTWVIDVDNVVKNITTESBIO	1363
Db	2035	-----PQTISTKPSYVDLKAQADENKITLSLLHSNDQIFEDKNNHLKIAKISIDELD	2090
QY	1364	KGELIENILS-----KLANIEDNKIILNNHTIN-----FYGDINE	1397
Db	2091	AHQIINSTTHDYDTLTKENNEWLLKTLQNLINKENTKYRVKVKWFTSKPSPDTIYGVNK	2150
QY	1398	SNRFISLTPSILEDINIIIEIDLVSYSKILLSGNCKMLIBNSSDIOOKID-HIGFN---	1453
Db	2151	DN-----IVYPI--NNHTNIDLKTEKS-----TLNSVKPTKTNFKNESIKLMVGNKGTG	2200
QY	1454	-GEHOKYIPYSIDNETKNGFIDYSKXEGFLPTABFSNESIIRNIYMPDSNNLFIYSKD	1512
Db	2201	SLENKYAKLVYKDNNNQIESDPVLLNSDTTTPKPFKEFSFN-----NQTSLVANRD	2254
QY	1513	LKDRII--NKGDVKL-----LIGNYFKDDMKVLSFTIEDTNTIKLNG-----VYL--	1557
Db	2255	FEFVKLIISDADFRKANLTDLTPNFNKDDAKFSIEPTPISVNNVIOGSNTYDEIHLSF	2314
QY	1558	-----DENGVAQ--ILKFMNNAKALNTSN---SLMNFLESINIKNIFVNNL-----	1599
Db	2315	DYDDQDHNLVNDQITITYRKKGQWTSKAGEVSVKNKINVLTKDLTPTTYEIGAI	2374
QY	1600	-----DPNIE-----FILDNTFIIGSNSIGOFELICD	1627
Db	2375	SITHVHSPNVKTSPIQYNETTKOLEKLDLTKVAHNFIDKFKLASNSKLVGVVE--CA	2432
QY	1628	KXKNTOPIYFNKI---KE-----TSYTLV-----GNRONLIVEPSYHLDDSGNISSTVIN	1676
Db	2433	KDENVK-----FSVGLHKEGHEGEYIYAKVDKTNPN---KPAIEIKSSNNINS---N	2481
QY	1677	FSQKLYGIDRYVNVKVIAPN	1697
Db	2482	SNQVYEFEL-----LKSQITPN	2498

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 16:37:55 / Search time 38.25 Seconds
(without alignments)
8413.819 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1020

Perfect score: 5189

Sequence: 1 MNLVNAQLQKVVYKFRIQ.....ITDASKVELVSTALDEITD 1020

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5189	100.0	2364	2 Q46342	Q46342 clostridium
2	4105.5	79.1	2367	2 Q9EXR0	Q9EXR0 clostridium
3	4096.5	78.9	2367	2 Q9F931	Q9F931 clostridium
4	4088.5	78.8	2367	2 Q46034	Q46034 clostridium
5	2156.5	41.6	554	2 Q93L39	Q93L39 clostridium
6	1916	36.9	698	2 Q68653	Q68653 clostridium
7	1898	36.6	698	2 Q9EXQ8	Q9EXQ8 clostridium
8	1892	36.5	697	2 Q86141	Q86141 clostridium
9	1474	28.4	553	2 Q9FCX5	Q9FCX5 clostridium
10	1435.5	27.7	2178	2 Q46149	Q46149 clostridium
11	422.5	8.1	3169	2 Q9ZGR4	Q9ZGR4 escherichia
12	419.5	8.1	3169	2 Q82916	Q82916 escherichia
13	391	7.5	3223	2 Q8L1L9	Q8L1L9 escherichia
14	389	7.5	3223	2 Q9RPH1	Q9RPH1 escherichia
15	386	7.4	3223	2 Q9RM48	Q9RM48 escherichia
16	386	7.4	3223	2 Q8VQR2	Q8VQR2 escherichia

17	382	7.4	3223	2 Q8VNR6	Q8VNR6 escherichia
18	358.5	6.9	2771	5 Q26216	Q26216 plasmodium
19	356	6.9	3346	16 Q822X0	Q822X0 chlamydomphi
20	352	6.8	3317	16 Q8EWP8	Q8EWP8 mycoplasma
21	350	6.7	3225	16 Q9PKM6	Q9PKM6 chlamydia m
22	348.5	6.7	2752	5 Q9BJY0	Q9BJY0 plasmodium
23	345.5	6.7	1127	12 Q9VYT6	Q9VYT6 melanoplus
24	345	6.6	5767	5 Q81525	Q81525 plasmodium
25	343	6.6	2965	5 Q8WRS5	Q8WRS5 plasmodium
26	342	6.6	2770	5 Q7YUE9	Q7YUE9 plasmodium
27	341	6.6	2976	5 Q8WRS4	Q8WRS4 plasmodium
28	338	6.5	2204	5 Q8IL70	Q8IL70 plasmodium
29	337.5	6.5	2867	5 Q9N2M3	Q9N2M3 plasmodium
30	336.5	6.5	2747	5 Q9BJX9	Q9BJX9 plasmodium
31	335.5	6.5	2957	5 Q8WRS6	Q8WRS6 plasmodium
32	328	6.3	3724	5 Q77320	Q77320 plasmodium
33	321.5	6.2	1238	12 Q9EMP3	Q9EMP3 amsecta moo
34	319.5	6.2	2166	16 O51465	O51465 borrelia bu
35	318.5	6.1	3394	5 Q77384	Q77384 plasmodium
36	317.5	6.1	3096	5 Q7YWF2	Q7YWF2 plasmodium
37	317.5	6.1	3254	5 Q9BK45	Q9BK45 plasmodium
38	317.5	6.1	3254	5 Q7YWE9	Q7YWE9 plasmodium
39	316.5	6.1	3081	5 Q7YWF1	Q7YWF1 plasmodium
40	316.5	6.1	3203	5 Q7YWE7	Q7YWE7 plasmodium
41	316.5	6.1	3203	5 Q7YWE6	Q7YWE6 plasmodium
42	316.5	6.1	3255	16 Q9PKM8	Q9PKM8 chlamydia m
43	316	6.1	1956	5 Q8IL11	Q8IL11 plasmodium
44	316	6.1	3256	5 Q7YWE8	Q7YWE8 plasmodium
45	315.5	6.1	1387	5 Q9GZ76	Q9GZ76 plasmodium

ALIGNMENTS

RESULT 1

Q46342 PRELIMINARY; PRT; 2364 AA.

AC Q46342; DT 01-NOV-1996 (TRENBLrel. 01, Created)

DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)

DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)

DE Cytotoxin L.

OS Clostridium sordellii.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI_TaxID=1505;

RN [1] SEQUENCE FROM N.A.

RC STRAIN=6018;

RA MEDLINE=95369733; PubMed=7642137;

RA Green G.A., Schue V., Montell H.;

RT "Cloning and characterization of the cytotoxin L-encoding gene of Clostridium sordellii: homology with Clostridium difficile cytotoxin B.";

RT B.";

RL Gene 161:57-61(1995).

RN [2] SEQUENCE FROM N.A.

RC STRAIN=6018;

RX MEDLINE=96149194; PubMed=9544213;

RA Green G.A., Schue V., Girardot R., Montell H.;

RT "Characterisation of an enterotoxin-negative, cytotoxin-positive strain of Clostridium sordellii.";

RT J. Med. Microbiol. 44:60-64(1996).

DR EMBL; X82638; CAA57959.1; -.

DR PIR; I40884; I40884.

DR GO; GO:0003743; F:translation initiation factor activity; IEA.

DR GO; GO:0006433; P:translational initiation; IEA.

DR InterPro; IPR002479; CW binding.

DR InterPro; IPR007577; Gly_transf_sug.

DR InterPro; IPR001950; TIF_SUI1.

DR Pfam; PF01473; CW_binding_1; 18.

DR Pfam; PF04488; Gly_transf_sug; 1.

DR PROSITE; PS01118; SUI1_1; 1.


```
SQ SEQUENCE 2364 AA; 270576 MW; EAD8A4467A89BDBB CRC64;

Query Match      100.0%; Score 5189; DB 2; Length 2364;
Best Local Similarity 100.0%; Pred. No. 3.4e-190;
Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNLVNAQLOKQVYKFRIOEDYVAIINALEEYHNMSSESVVEKYKLKDINNLTNYL 60
DB 1 MNLVNAQLOKQVYKFRIOEDYVAIINALEEYHNMSSESVVEKYKLKDINNLTNYL 60

QY 61 NTYKSGRNKALKKFKKEYLTMEVLEKNSLTPVEKNLHF*WIGQINDTAINYNQWKD 120
DB 61 NTYKSGRNKALKKFKKEYLTMEVLEKNSLTPVEKNLHF*WIGQINDTAINYNQWKD 120

QY 121 VNSDYTVKVFYDSNAPLINTLKTIVESATNTLTESFRENLDPEFDYKFKRMEIYY 180
DB 121 VNSDYTVKVFYDSNAPLINTLKTIVESATNTLTESFRENLDPEFDYKFKRMEIYY 180

QY 181 DKQKHFDYKSGQIENPEFIIDNTIKTYLSNEYSKOLEALNKYTESLNKITTANNNDI 240
DB 181 DKQKHFDYKSGQIENPEFIIDNTIKTYLSNEYSKOLEALNKYTESLNKITTANNNDI 240

QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDIIRISMLKEDGGVLDVDILPGIOPDLFKS 300
DB 241 RNLEKFADEDLVRLYNQELVERWNLAAASDIIRISMLKEDGGVLDVDILPGIOPDLFKS 300

QY 301 INKPDSITNTSWEMIKLEAIIKMYKEYIPGYT*SKNFDMLDEEVQVSFESALSSKSDKSEIF 360
DB 301 INKPDSITNTSWEMIKLEAIIKMYKEYIPGYT*SKNFDMLDEEVQVSFESALSSKSDKSEIF 360

QY 361 LPLDDIKVSPLEVKIAFANNVINOALISLKDSCSYDLVINQIKRKYKILNDNLNPSINE 420
DB 361 LPLDDIKVSPLEVKIAFANNVINOALISLKDSCSYDLVINQIKRKYKILNDNLNPSINE 420

QY 421 GTDFNTWKIPSDKLIASISNEDNMFMFKITNYLKVGFAPDVRSTINLSGCVVTGYAQD 480
DB 421 GTDFNTWKIPSDKLIASISNEDNMFMFKITNYLKVGFAPDVRSTINLSGCVVTGYAQD 480

QY 481 LLMFKDNSTNIHLLPELNEFPFKTKISQTEQBITSLWFSFNQARAKSPPEYKKGYPE 540
DB 481 LLMFKDNSTNIHLLPELNEFPFKTKISQTEQBITSLWFSFNQARAKSPPEYKKGYPE 540

QY 541 GALGEDNDLPAQNTVLDKDYVSKKILSSMKTRNKEYIHYIVQOGDKISYEASCNLFSSK 600
DB 541 GALGEDNDLPAQNTVLDKDYVSKKILSSMKTRNKEYIHYIVQOGDKISYEASCNLFSSK 600

QY 601 DPYSILYOKNIEGSETAYYYVADAETKEIDKYRIPYCI*SNKNIKLTFIGHGKSEFNT 660
DB 601 DPYSILYOKNIEGSETAYYYVADAETKEIDKYRIPYCI*SNKNIKLTFIGHGKSEFNT 660

QY 661 DTFANLVDVLSLSEETITINLAKADISPKYIENLLGCNMFYSISAEETYPGKLLKIK 720
DB 661 DTFANLVDVLSLSEETITINLAKADISPKYIENLLGCNMFYSISAEETYPGKLLKIK 720

QY 721 DRVSELMPISQDSITVSANQVEYRINPEGKEIILDHSGKWINKEESIIKDISKEYISF 780
DB 721 DRVSELMPISQDSITVSANQVEYRINPEGKEIILDHSGKWINKEESIIKDISKEYISF 780

QY 781 NPENKIIYKSKYLHELSTLLOEIRNANSDDIDLEKKVMT*TECEINVASNIDRQIVEGR 840
DB 781 NPENKIIYKSKYLHELSTLLOEIRNANSDDIDLEKKVMT*TECEINVASNIDRQIVEGR 840

QY 841 IBEAKNLTSDSINYIKNEFKL*ESTSDSLYDLKHONGLDDSHFISFEDISTENGFRIRF 900
DB 841 IBEAKNLTSDSINYIKNEFKL*ESTSDSLYDLKHONGLDDSHFISFEDISTENGFRIRF 900

QY 901 INKETGNSIFIETEKEIFESEYATHISKEISNIKOTIFDNVNGKLVKYNLDAAEVNTLN 960
DB 901 INKETGNSIFIETEKEIFESEYATHISKEISNIKOTIFDNVNGKLVKYNLDAAEVNTLN 960

QY 961 SAFFIQSLIEYNTTKESLNLSVAMKVQVYAOQFSTGLNTITDASKVVELVSTALDEIID 1020
DB 961 SAFFIQSLIEYNTTKESLNLSVAMKVQVYAOQFSTGLNTITDASKVVELVSTALDEIID 1020
```

RESULT 2

```
Q9EXR0 PRELIMINARY; PRT: 2367 AA.
AC Q9EXR0;
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Toxin B.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8864;
RA von Bichel-Streiber C.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=8864;
RA Kohl M.;
RT "The pathogenicity locus of Clostridium difficile, strain 8864.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011301; CAC19891.1; -.
DR PIR; A27636; A27636.
DR GO; GO:0003743; F:translation initiation factor activity; IEA.
DR GO; GO:0006413; P:translational initiation; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR007577; Gly_transf_sug.
DR InterPro; IPR001950; TIF_SUI1.
DR Pfam; PF01473; CW_binding 1; 16.
DR Pfam; PF04488; Gly_transf_sug; 1.
DR PROSITE; PS01118; SUI1_1_1.
SQ SEQUENCE 2367 AA; 269072 MW; BF7087C92E7389E7 CRC64;
```

Query Match 79.1%; Score 4105.5; DB 2; Length 2367;
Best Local Similarity 77.4%; Pred. No. 8e-149;
Matches 790; Conservative 114; Mismatches 116; Indels 1; Gaps 1;

```
QY 1 MNLVNAQLOKQVYKFRIOEDYVAIINALEEYHNMSSESVVEKYKLKDINNLTNYL 60
DB 1 MSLVNRKQLEKAVNRFRVQEDYVAIDALEEYHNMSSENTVVEKYKLKDINSLTDYI 60

QY 61 NTYKSGRNKALKKFKKEYLTMEVLEKNSLTPVEKNLHF*WIGQINDTAINYNQWKD 120
DB 61 DTYKSGRNKALKKFKKEYLTVEILEKNSLTPVEKNLHF*WIGQINDTAINYNQWKD 120

QY 121 VNSDYTVKVFYDSNAPLINTLKTIVESATNTLTESFRENLDPEFDYKFKRMEIYY 180
DB 121 VNSDYTVKVFYDSNAPLINTLKTIVESATNTLTESFRENLDPEFDYKFKRMEIYY 180

QY 181 DKQKHFDYKSGQIENPEFIIDNTIKTYLSNEYSKOLEALNKYTESLNKITTANNNDI 240
DB 181 DKQKHFDYKSGQIENPEFIIDNTIKTYLSNEYSKOLEALNKYTESLNKITTANNNDI 240

QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDIIRISMLKEDGGVLDVDILPGIOPDLFKS 300
DB 241 RNFEFEKTEGVNLYEQELVERWNLAAASDIIRISMLKEDGGVLDVDILPGIOPDLFKD 300

QY 301 INKPDSI-TNTSWEMIKLEAIIKMYKEYIPGYT*SKNFDMLDEEVQVSFESALSSKSDKSEI 359
DB 301 INKPDSVKYAVDWEENQLEAIIKMYKEYIPGYT*SKNFDMLDEEVQVSFESALSSKSDKSEI 360

QY 360 FLPLDDIKVSPLEVKIAFANNVINOALISLKDSCSYDLVINQIKRKYKILNDNLNPSIN 419
DB 361 FLPLGGIEVSPLEVKVAPAKGSIIDQALISAKDSCSYDLVINQIKRKYKILNDNLNPSIN 420

QY 420 EGTDFNTWKIPSDKLIASISNEDNMFMFKITNYLKVGFAPDVRSTINLSGCVVTGYAQ 479
DB 421 QGDNFTTWNFGESLGAIANENISFIAKISYLRVGFYPEANTTITLSGPTIYAGAYK 480
```

QY 480 DLLMFKDNTNHLPELNPFPKTKISOLTEQBITSLWSPNQARAKSQEYKKGYP 539
 DB 481 DLLTFKMSIDTSLSELNFPFPKNTSQATEQKNSLWQNEERAKIQEYKKNYP 540
 QY 540 EGALGEDNDLFAQNTVLDKDYYSKTLSSMKTRKEYIHYIVLOQDKISYEASCNLPS 599
 DB 541 EGALGEDNDLFSQNTVTDKEYLLEKISSKTSSEYVHYIVLOQDKISYEACNLPA 600
 QY 600 KDPYSSILYQKNEGSETATYYVADAEIKEDIKYRIPYQISNKNKILTFIGHGKSEFN 659
 DB 601 KNPYDSILFQKNEDEVAYYNPTDSEIOEDKYPIDRISDRPKIKLTLIGHGKAEFN 660
 QY 660 TDTFANLDVDSLSEIETILNLAKADISPKYIIEINLLGCMFYSISABETYPGKLLKI 719
 DB 661 TDTFAGLDVDSLSEIETILDLAKADISPKSIEINLLGCMFYSYNVVEETYPGKLLRV 720
 QY 720 KDRVSELMPSISQDSITVSANOYEVINEEGKEBEILDHSGKWINKESIIKOISSKEYIS 779
 DB 721 KDKVSELMPSISQDSIIVSANQYEVINSEGRRELLDHSGEWINKESIIKOISSKEYIS 780
 QY 780 FNPENKIIIVKSKYLHELSTLLOEIRNANSDDILEKKVMLTECEINVASNIDROIVEG 839
 DB 781 FNPENKIIIVKSKYLPSTLLQEIERNNSDDIELEKVMLAECIEINVISNIEQVVEE 840
 QY 840 RIEBAKNTLSDSINYIKNEFKLIESISDSLYDLKHQNGLDDSHFISFEDISKTEGFRIR 899
 DB 841 RIEBAKSLTSDSINYIKNEFKLIESISDALYDLKQNELEESHFISFEDISKTEGFSIR 900
 QY 900 FINKETGNSIFETEKEIFSEYATHISKEISNIDKTIIPDNVNGKLVKKNLDAAHEVNTL 959
 DB 901 FIDKETGESIFVETEKAFSEYANHITTEISKLKDTIFDVTNGKLVKKTLDATHEVNTL 960
 QY 960 NSAFPIQSILIEYNTTKESLSNLSVAMKVQVYAOLFSTGLNTITDASKVVELVSTALDETI 1019
 DB 961 NAAFTQSLIGYSSKESLSNLSVAMKVQVYAOLFSTGLNTITDAAKVVVELVSTALDETI 1020
 QY 1020 D 1020
 DB 1021 D 1021
 RESULT 3
 Q9F931 PRELIMINARY; PRT: 2367 AA.
 AC Q9F931;
 DT 01-MAR-2001 (TremBrel. 16, Created)
 DT 01-MAR-2001 (TremBrel. 16, Last sequence update)
 DT 01-JUN-2003 (TremBrel. 24, Last annotation update)
 DE Cytotoxin B.
 GN TCDB.
 OS Clostridium difficile.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_taxID=1496;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=5340;
 RX MEDLINE=20448897; PubMed=10992443;
 RA Sambol S.P., Merrigan M.M., Lively D., Gerdling D.N., Johnson S.;
 RT "Toxin gene analysis of a variant strain of clostridium difficile that
 RT causes human clinical disease";
 RL Infect. Immun. 68:5480-5487(2000).
 DR EMBL; AF217292; AAG18011.1; -
 DR PIR; A27636; A27636.
 DR GO; GO:0003743; P:translational initiation factor activity; IEA.
 DR GO; GO:0006413; P:translational initiation; IEA.
 DR InterPro; IPR002479; CW binding.
 DR InterPro; IPR007577; Gly transf_sug.
 DR InterPro; IPR001950; Tif SUI1.
 DR Pfam; PF01473; CW binding_1; 17.
 DR Pfam; PF04488; Gly transf_sug; 1.
 DR PROSITE; PS01118; SUI1_1; 1.

SQ SEQUENCE 2367 AA; 269337 MW; D5BE715ESBD41E2F CRC64;
 Query Match 78.9%; Score 4096.5; DB 2; Length 2367;
 Best Local Similarity 77.0%; Pred. No. 1.8e-148;
 Matches 786; Conservative 117; Mismatches 117; Indels 1; Gaps 1;
 QY 1 MNLVKAQLOQVWVVKRIQEDVVAIINALVEYHNHNSSEVVEKYKLKDINLNTDNYL 60
 DB 1 MSLVNRKQLEKQANRFRVQEDVVAIDALVEYHNHNSSEVVEKYKLKDINLNTDNYL 60
 QY 61 NTKKSGRNKALKFKFKEYLTMEVLELKNLSLTPVEKQLHFIIWIGQINDTAINYNQWKD 120
 DB 61 DTYKSGRNKALKFKFKEYLTMEVLELKNLSLTPVEKQLHFIIWIGQINDTAINYNQWKD 120
 QY 121 VNSDYTVKVFYDSNAFLINTLTKTIVESATNTLTSPRENLDPEFDYKPYKRMELIY 180
 DB 121 VNSDYTVKVFYDSNAFLINTLTKTIVESATNTLTSPRENLDPEFDYKPYKRMELIY 180
 QY 181 DKQKHFDIYKSOEENPEFIIDNIIKTYLSNEYSKOLEALNKYIEESLNKITANNNDI 240
 DB 181 DKQKFNINYYKAAQKEENPDIIIDIVKTYLSNEYSKODIDELNAYIEESLNKVTENSNDV 240
 QY 241 RNLEKFADEDLVRLYNOELVERWNLAAASDILRISMLKEDGGVYLDVILPGIOPDLFKS 300
 DB 241 RNFBEFTGEVFNLYEQLVERWNLAGASDILRVAILKXNIGVYLDVDMLEPSIHPDLFKD 300
 QY 301 INKPDISI-TNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQSPESALSSKSDSEI 359
 DB 301 INKPDVSVKTAVDWEEMQLEAIMKHKEYIPEYTSKHFDTLDEEVQSSPESVASKSDSEI 360
 QY 360 FLPLDDIKVPLEVKIAFANNVINOALISKDSYCSDLVINQIKRYKILINDNLNPSIN 419
 DB 361 FLPLGDIKVSLEVKIAFAGKSIINOALISAKDSYCSDDLKQIKQNYKIINDLGLPIS 420
 QY 420 EGTDPNTMTKIFSKLASINEDNMWMIKITNLYKVPADPDRVSTINLSPGVYTGAYQ 479
 DB 421 QGNDTNTMNFSGESLGAIAENENISFTAKIGSVLRGVFPYEAINTVTLSTGPTTYAGAYK 480
 QY 480 DLLMFKDNTNHLPELNPFPKTKISOLTEQBITSLWSPNQARAKSQEYKKGYP 539
 DB 481 DLLTFKMSIDTSLSELNFPFPKNTSQATEQKNSLWQNEERAKIQEYKKNYP 540
 QY 540 EGALGEDNDLFAQNTVLDKDYVSKKLSSMKTRKEYIHYIVLOQDKISYEASCNLPS 599
 DB 541 EGALGEDNDLFSQNTVTDKEYLLEKISSKTSSEYVHYIVLOQDKISYEACNLPA 600
 QY 600 KDPYSSILYQKNEGSETATYYVADAEIKEDIKYRIPYQISNKNKILTFIGHGKSEFN 659
 DB 601 KNPYDSILFQKNEDEVAYYNPTDSEIOEDKYPIDRISDRPKIKLTLIGHGKAEFN 660
 QY 660 TDTFANLDVDSLSEIETILNLAKADISPKYIIEINLLGCMFYSISABETYPGKLLKI 719
 DB 661 TDTFAGLDVDSLSEIETIAGLAKEDISPKSIEINLLGCMFYSYNVVEETYPGKLLRV 720
 QY 720 KDRVSELMPSISQDSITVSANOYEVINEEGKEBEILDHSGKWINKESIIKOISSKEYIS 779
 DB 721 KDKVSELMPSISQDSIIVSANQYEVINSEGRRELLDHSGEWINKESIIKOISSKEYIS 780
 QY 780 FNPENKIIIVKSKYLHELSTLLOEIRNANSDDILEKKVMLTECEINVASNIDROIVEG 839
 DB 781 FNPENKIIIVKSKYLPSTLLQEIERNNSDDIELEKVMLAECIEINVISNIEQVVEE 840
 QY 840 RIEBAKNTLSDSINYIKNEFKLIESISDSLYDLKHQNGLDDSHFISFEDISKTEGFRIR 899
 DB 841 RIEBAKSLTSDSINYIKNEFKLIESISDALCQNELEDSHFISFEDISDEGFSIR 900
 QY 900 FINKETGNSIFETEKEIFSEYATHISKEISNIDKTIIPDNVNGKLVKKNLDAAHEVNTL 959
 DB 901 FINKETGESIFVETEKTIIFSEYANHITTEISKIKOTIIFDVTNGKLVKKNLDTHEVNTL 960
 QY 960 NSAFPIQSILIEYNTTKESLSNLSVAMKVQVYAOLFSTGLNTITDASKVVELVSTALDETI 1019
 DB 961 NAAFTQSLIGYSSKESLSNLSVAMKVQVYAOLFSTGLNTITDAAKVVVELVSTALDETI 1020

```

QY 1020 D 1020
Db 1021 D 1021

RESULT 4
Q46034 PRELIMINARY; PRT; 2367 AA.
AC Q46034;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Toxin B.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
ON NCBI_TaxID=1496;
RX STRAIN=96079281; PubMed=7494480;
RA von Eichel-Streiber C., Meyer zu Heringdorf D., Habermann E.,
RA Saringen S.;
RT "Closing in on the toxic domain through analysis of a variant
RT Clostridium difficile cytotoxin B.";
RL Mol. Microbiol. 17:313-321(1995).
DR EMBL; Z23277; CAA80815.1; -.
DR PIR; A27636; A27636.
DR GO; GO:0006413; P:translational initiation; IEA.
DR GO; GO:003743; F:translation initiation factor activity; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR001950; TIF SUI1.
DR Pfam; PF01473; CW binding 1; 17.
DR Pfam; PF04488; Gly transf_sug; 1.
DR PROSITE; PS01118; SUI1_1; 1.
SQ SEQUENCE 2367 AA; 269186 MW; BF9823DAE70427F3 CRC64;

Query Match 78.8%; Score 4088.5; DB 2; Length 2367;
Best Local Similarity 77.1%; Pred. No. 3.6e-148;
Matches 787; Conservative 114; Mismatches 119; Indels 1; Gaps 1;

QY 1 MSLVNRKALQKQVYKFRQEDYVAILNALBEYHNMSESSVVEKYKLKQINNTDNYL 60
Db 1 MSLVNRKQLEKQVNRVQEDYVAILDALBEYHNMSESSVVEKYKLKQINSLTDYI 60

QY 61 NTKYKGRKALKKPKKYLTMVELEKNSLTPVEKNLHFIWIGQINDTAINYNQWKD 120
Db 61 DTYKSGRNLKALKKPKKYLTMVELEKNSLTPVEKNLHFIWIGQINDTAINYNQWKD 120

QY 121 VNSDYFVKVYDSNAFLNTLTKTIVESATNTLSEFRENLDNPEFDYKFKRMEIY 180
Db 121 VNSDYNVNFYDSNAFLNTLTKTIVESATNTLSEFRENLDNPEFDYKFKRMEIY 180

QY 181 DKQKHFIDYKSGIEENPFIIIDNIKTYSNYSKDEALNKYIEESLNKITANNNDI 240
Db 181 DKQKHFIDYKSGIEENPFIIIDNIKTYSNYSKDEALNKYIEESLNKITANNNDI 240

QY 241 RNLEKPADBDLVNLQELVERWNLAASDIIRISMLKEDGGVYLDVILPGIQDPLKS 300
Db 241 RNLEKPADBDLVNLQELVERWNLAASDIIRISMLKEDGGVYLDVILPGIQDPLKS 300

QY 301 INKPSDI-TNWSNEMIKLAIWKYIPIGYTSKNFDMLEDEVSFESALSSDKSEI 359
Db 301 INKPSDI-TNWSNEMIKLAIWKYIPIGYTSKNFDMLEDEVSFESALSSDKSEI 359

QY 360 FLPLDIDKVSPLKVKIAFANNVSNQALISLQKSDYSLVINQIKNRYKILNDNLNPSIN 419
Db 360 FLPLDIDKVSPLKVKIAFANNVSNQALISLQKSDYSLVINQIKNRYKILNDNLNPSIN 419

QY 361 FLPLGDIKVSPLKVKIAFANNVSNQALISLQKSDYSLVINQIKNRYKILNDNLNPSIN 420
Db 361 FLPLGDIKVSPLKVKIAFANNVSNQALISLQKSDYSLVINQIKNRYKILNDNLNPSIN 420

QY 420 EGTDFNTWKIFSDKILASINEDNMFMKIKTNLYKVGFPAPDVRSTINLSGPGVYTGAYQ 479
Db 420 EGTDFNTWKIFSDKILASINEDNMFMKIKTNLYKVGFPAPDVRSTINLSGPGVYTGAYQ 479

Db 421 QGDNFNTMNFESIGALAEENISFIAGKSYLRVGFPEANTTITLSGFTIYAGAYK 480
QY 480 DILLMFKDNSTNIHLLEPELNRNFEPPKTKISQTEOBTISLWGFNOARAKSQEYKGYF 539
Db 481 DLTTFKMSIDTISLSSELNRNFEPPKTKISQTEOBTISLWGFNOARAKSQEYKGYF 540
QY 540 EGALGEDDNLDFQNTVLDKDYVSKKILSSMKTRNKYIHYIVQLQGGKISYEASCNLF 599
Db 541 EGALGEDDNLDFQNTVLDKDYVSKKILSSMKTRNKYIHYIVQLQGGKISYEASCNLF 600
QY 600 KDPYSILYQKNIQEGSETAYVYVADAETKEIDKRYIPYQISNKNKNIKUTFTGHGKSE 659
Db 601 KNPYDSILFQNTIEDSEVAYVYVADAETKEIDKRYIPYQISNKNKNIKUTFTGHGKSE 660
QY 660 TDTFANLVDVSLSSSEIETTLNLAADISPKYIEINLLGCNMFYSYSABETYPGKLLKI 719
Db 661 TDTFANLVDVSLSSSEIETTLNLAADISPKYIEINLLGCNMFYSYSABETYPGKLLKI 720
QY 720 KDRVSELMPSISODSITVSANOVYEVNRINBEGKREILDHSGKWKINKEESIKDISKEYIS 779
Db 721 KDRVSELMPSISODSITVSANOVYEVNRINBEGKREILDHSGKWKINKEESIKDISKEYIS 780
QY 780 FNPENKNIIVKSKYLHELSTLQEIIRNNANSDDIDLEKKVMLTECEINVASNIDRIVEG 839
Db 781 FNPENKNIIVKSKYLHELSTLQEIIRNNANSDDIDLEKKVMLTECEINVASNIDRIVEG 840
QY 840 RIBEAKNLTSDSINYIKNEFKLIESDSLYDLKHONGLDHSHFISFEDISTENGFRIR 899
Db 841 RIBEAKNLTSDSINYIKNEFKLIESDSLYDLKHONGLDHSHFISFEDISTENGFRIR 900
QY 900 FINKETGNSIFIEKEIFESEVATHISKEISNKTDFPNVNGKLVKYNLDAAEHVNTL 959
Db 901 FINKETGNSIFIEKEIFESEVATHISKEISNKTDFPNVNGKLVKYNLDAAEHVNTL 960
QY 960 NSAFFIQSLIEYNTTKESLNLSVAMKVQVYQALFSTGLNTITDASKVVELVSTALDETI 1019
Db 961 NSAFFIQSLIEYNTTKESLNLSVAMKVQVYQALFSTGLNTITDASKVVELVSTALDETI 1020

QY 1020 D 1020
Db 1021 D 1021

RESULT 5
Q93139 PRELIMINARY; PRT; 554 AA.
AC Q93139;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE TCD-B-C34 Cluster1-2 (Fragment).
GN TCD-B-C34.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
ON NCBI_TaxID=1496;
RX STRAIN=C34;
DT MEDLINE=2133121; PubMed=11430410;
RA Mehlig M., Moos M., Braun V., Kalt B., Mahony D.E.,
RA von Eichel-Streiber C.;
RT "Variant toxin B and a functional toxin A produced by Clostridium
RT difficile C34.";
RL FEMS Microbiol. Lett. 198:171-176(2001).
DR EMBL; AJ294944; CAC41640.1; -.
DR InterPro; IPR007577; Gly transf_sug.
DR Pfam; PF04488; Gly transf_sug; 1.
FT NON TER 554
SQ SEQUENCE 554 AA; 63910 MW; 9285512AC1B1D87A CRC64;

Query Match 41.6%; Score 2156.5; DB 2; Length 554;

```

```

Best Local Similarity 73.6%; Pred. No. 4.4e-75;
Matches 408; Conservative 67; Mismatches 78; Indels 1; Gaps 1;

QY 1 MNLVNAQLOKQVYKFRIOEDVVAIINALAEYHNMSSESVVEKYKLKDINNLTNYL 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MSLISKEBELIKLAY-SIRPRENEYKVTLTNLDYKNTNNENKYLQKKLNEIDVFM 59

QY 61 NTKKSGRNKALKKFEKYLTVLELKNNSLTPVEKNLHFWIGQINDTAINYQWKD 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 60 NTKKSGRNKALKKFEKYLTVLELKNNSLTPVEKNLHFWIGQINDTAINYQWKD 119

QY 121 VNSDYTVKVFYDSNAFLINTLTKTIVBSATNTTLESFRENLDPEFDYKFKRMEIY 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 INAEYTNKLWYDSEAFVNTLTKALVESSTTEALQLLEEIQNPOFDMNKYKRMETIY 179

QY 181 DQKHFDIYKSOQEBENPEFIIDNIITKYLSEYKSKOLEALNKYTESLNKITANGNDI 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 180 DQKHFDIYKSOQEBENPEFIIDNIITKYLSEYKSKOLEALNKYTESLNKITANGNDI 239

QY 241 RNLEKPADEDLVRLYNQBELVERWNLAAASDLIRLSMLKEDGGVYLDVILPGIOPDLFKS 300
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 240 RANSLFTEQELNINYNQBELVERWNLAAASDLIRLSMLKEDGGVYLDVILPGIOPDLFKT 299

QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQSFESALSKSKSIF 360
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 300 ISRPSISGLDRWEMIKLEAIMKYKEYINNYTSENPDQLDQOLKONFKLIIESKSEKSEIF 359

QY 361 LPLDDIKVSPLEVKIAFANNVINOALISLSDSYCSDLVINOIKNRYKILNDNLNPSINE 420
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 360 SKLENLWSDLEIKIAPALGVSINOALISLSDSYCSDLVINOIKNRYKILNDNLNPSINE 419

QY 421 GTDFNTMTKIFSDKLASINEDNMFMKITNYLKVGFAPDVRSTINLSGPGVYTCAYQD 480
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 420 DNNFTDTTKIFHDSLFNSATAENSMLTKIAPYLQVGFMPPEARSTISLSPGAYASAYD 479

QY 481 LLMFKDNTNHLLEPELRFNPEFPKTKISOLTEQBITSLWSFNQARAKSQBEYKKGYFE 540
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 480 FNLQENTIEKTKASDLIEFKFPENNLSOLTEQBITSLWSFNQARAKSQBEYKKGYFE 539

QY 541 GALGEDDNLDFQAQTVLDKDY-VSKILSS--MKTRNKEYIHYIVQLQGDKITSEASCNL 597
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 540 GSLSDNEVDNFKNLTALDKNLLNKNIPSNVVEAGSKNYHYIIQLQGDKITSEATCNL 599

QY 598 FSKDEYSSILYOKNTEGSETAYVYVAD--ABEIKDKYRIPYQISNKNRINKLTFIGRKG 655
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 600 FSKNPKNSIIIORNM--NESAKSYFLSDGDSILELKNRYPERLKNKEKVKVTFIGHGK 657

QY 656 SEFNTDTTANLDVDSLSSEITILNLAKADISPKYIEINLL 696
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 658 DEFNTSEFARLSVDSLSSEISFLDTIKLDISPKXNVEVNLL 698

RESULT 7
Q9EXQ8 PRELIMINARY; PRT; 698 AA.
AC Q9EXQ8;
DT 01-VAR-2001 (TremBLrel. 16, Created)
DT 01-VAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Truncated toxin A.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8864;
RA von Eichel-Streiber C.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=8864;
RA Kohl M.;
RT "The pathogenicity locus of Clostridium difficile, strain 8864."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

Best Local Similarity 73.6%; Pred. No. 4.4e-75;
Matches 408; Conservative 67; Mismatches 78; Indels 1; Gaps 1;

QY 1 MNLVNAQLOKQVYKFRIOEDVVAIINALAEYHNMSSESVVEKYKLKDINNLTNYL 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MSLISKEBELIKLAY-SIRPRENEYKVTLTNLDYKNTNNENKYLQKKLNEIDVFM 59

QY 61 NTKKSGRNKALKKFEKYLTVLELKNNSLTPVEKNLHFWIGQINDTAINYQWKD 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 60 NTKKSGRNKALKKFEKYLTVLELKNNSLTPVEKNLHFWIGQINDTAINYQWKD 119

QY 121 VNSDYTVKVFYDSNAFLINTLTKTIVBSATNTTLESFRENLDPEFDYKFKRMEIY 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 INAEYTNKLWYDSEAFVNTLTKALVESSTTEALQLLEEIQNPOFDMNKYKRMETIY 179

QY 181 DQKHFDIYKSOQEBENPEFIIDNIITKYLSEYKSKOLEALNKYTESLNKITANGNDI 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 180 DQKHFDIYKSOQEBENPEFIIDNIITKYLSEYKSKOLEALNKYTESLNKITANGNDI 239

QY 241 RNLEKPADEDLVRLYNQBELVERWNLAAASDLIRLSMLKEDGGVYLDVILPGIOPDLFKS 300
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 240 RANSLFTEQELNINYNQBELVERWNLAAASDLIRLSMLKEDGGVYLDVILPGIOPDLFKT 299

QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQSFESALSKSKSIF 360
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 300 ISRPSISGLDRWEMIKLEAIMKYKEYINNYTSENPDQLDQOLKONFKLIIESKSEKSEIF 359

QY 361 LPLDDIKVSPLEVKIAFANNVINOALISLSDSYCSDLVINOIKNRYKILNDNLNPSINE 420
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 360 SKLENLWSDLEIKIAPALGVSINOALISLSDSYCSDLVINOIKNRYKILNDNLNPSINE 419

QY 421 GTDFNTMTKIFSDKLASINEDNMFMKITNYLKVGFAPDVRSTINLSGPGVYTCAYQD 480
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 420 DNNFTDTTKIFHDSLFNSATAENSMLTKIAPYLQVGFMPPEARSTISLSPGAYASAYD 479

QY 481 LLMFKDNTNHLLEPELRFNPEFPKTKISOLTEQBITSLWSFNQARAKSQBEYKKGYFE 540
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 480 FNLQENTIEKTKASDLIEFKFPENNLSOLTEQBITSLWSFNQARAKSQBEYKKGYFE 539

QY 541 GALGEDDNLDFQAQTVLDKDY-VSKILSS--MKTRNKEYIHYIVQLQGDKITSEASCNL 597
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 540 GSLSDNEVDNFKNLTALDKNLLNKNIPSNVVEAGSKNYHYIIQLQGDKITSEATCNL 599

QY 598 FSKDEYSSILYOKNTEGSETAYVYVAD--ABEIKDKYRIPYQISNKNRINKLTFIGRKG 655
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 600 FSKNPKNSIIIORNM--NESAKSYFLSDGDSILELKNRYPERLKNKEKVKVTFIGHGK 657

QY 656 SEFNTDTTANLDVDSLSSEITILNLAKADISPKYIEINLL 696
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 658 DEFNTSEFARLSVDSLSSEISFLDTIKLDISPKXNVEVNLL 698

RESULT 6
O68653 PRELIMINARY; PRT; 698 AA.
AC O68653;
DT 01-AUG-1998 (TremBLrel. 07, Created)
DT 01-AUG-1998 (TremBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Truncated toxin A.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CCUG 20309;
RA MEDLINE=99304497; PubMed=10376209;
RA Song X.P., Bai X.L., Chang S.Y.;
RT "Nucleotide and peptide sequences of the open reading frame encoding a truncated toxin A gene of Clostridium difficile strain CCUG 20309."
RL DNA Seq. 10:93-96(1999).
DR EMBL; AF053400; AAC08437.1;
DR InterPro; IPR007577; Glytransf_sug.
DR Pfam; PF04488; Glytransf_sug; 1.
SQ SEQUENCE 698 AA; 80682 MW; 021EB268A3BDEC5E CRC64;

Query Match 36.9%; Score 1916; DB 2; Length 698;
Best Local Similarity 52.9%; Pred. No. 8.6e-66;
Matches 371; Conservative 138; Mismatches 184; Indels 8; Gaps 5;

```



```
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Toxin A (Fragment).
GN TCDA.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C34;
RX MEDLINE=20402122; PubMed=10931284;
RA Braun V., Mehlig M., Moos M., Rupnik M., Kalt B., Mahony D.,
RA von Eichel-Streiber C.;
RT "A chimeric ribozyme in Clostridium difficile combines features of
RT group I introns and insertion elements.";
RL Mol. Microbiol. 36:1447-1459(2000).
DR EMBL; AJ131844; CAC03681.1; -.
DR InterPro; IPR007577; Gly_transf_sug.
DR Pfam; PF04488; Gly_transf_sug; 1.
FT NON TPR 553_553
SQ SEQUENCE 553 AA; 63913 MW; DDE4551A6D8C3B25 CRC64;

Query Match      28.4%; Score 1474; DB 2; Length 553;
Best Local Similarity 51.7%; Pred. No. 5.1e-49;
Matches 287; Conservative 108; Mismatches 158; Indels 2; Gaps 2;

QY 1 MNLVWKAQLOQMYVVKRFQDEDEVAAILNALBSEYHNMSESSVVEKYLKDKDINLTDNYL 60
DB 1 MSSISKELKLAY-SVRPRENEYKTLTNDYENKLTNNNNENKYLQKLKLNESIDVFM 59

QY 61 NTVKSGRNKALKKFKYLTWELKNNSLTPVEKNLHFIWIGGOINDTAINYINQWDX 120
DB 60 NKNYNSNRNALSNLKKDILKEVILKNSNTSPVEKNLHFWIGGEVSDIALEYIKQWAD 119

QY 121 VNSDYTVKVPYDSNAPLINTLTKTIVESATNNTLESFRENLPDPDYKFKRMEIYY 180
DB 120 INAEYVNLWYDSAPLVNTLKKALVESSSTTEALQLLEBEIQNPQDM-KFYKRWEIFY 178

QY 181 DKQGHFDYKSOIEENPEFIIDNIKTLSNEYSKDLEALNKYIEESLNKTANNNGDI 240
DB 179 DRQRRFINYKQINKPTVPTIDDIKSHVSEYNRDETLLSEYRTNSLRKINSNHGIDI 238

QY 241 RNLKEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVLDILPGIQPLFKS 300
DB 239 RANSLFTEQELNLYSQELNLRGTLAAASDIVALLAKNFGVYLDVLDILPGIHSDFKT 298

QY 301 INKPDSTINTSWEMIKLEAMKYKEYIPGYTSKNFMDLDEVORSFESALSSKSKSEIF 360
DB 299 IPRSSSIGLDRWEMIKLEAMKYKYINNTYSENFDKLDQQLKONFKLIESKSEIF 358

QY 361 LPLDDIKVSPLEVKIAPANNVINOALISLKDSYCSDLVINOIKNRYKIINDNLNPSINE 420
DB 359 SKLENLVSDLEIKIAPALGSVINOALISKQGSYLTNLVIEQVKRYQFNLQHLNPAIES 418

QY 421 GTDFTMTKFSDKLASISNEDNMFMKIITNYLKVPAPDVSTINLSGPGYVGTAYQD 480
DB 419 DNNFTDTTKIFHDSLFNSATAENSMFLTKIAPLVQGFMPPEARSTISLSGPGYASAYD 478

QY 481 LLMFKNSTNIHLEPELRPFPPKTKISOLTEOITSLWSFNOARAKSOFEYKKGYPE 540
DB 479 FINQENIETKLASDLIEBFKFPENNLSQTEOINLSWFSFQASAKYQIERYVDYTG 538

QY 541 GALGEDNLDFAQNT 555
DB 539 GSPSGDNGVDNFKNT 553

RESULT 10
ID Q46149 PRELIMINARY; PRT; 2178 AA.
AC Q46149; Q46147; Q46148;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Alpha-toxin.
OS Clostridium novyi.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1542;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC19402;
RX MEDLINE=95342160; PubMed=7616958;
RA Hofmann F., Herrmann A., Habermann E., von Eichel-Streiber C.;
RT "Sequencing and analysis of the gene encoding the alpha-toxin of
RT Clostridium novyi proves its homology to toxins A and B of Clostridium
RT difficile.";
RL Mol. Gen. Genet. 247:670-679(1995).
RN [2]
RP SEQUENCE OF 1204-2178 FROM N.A.
RC STRAIN=ATCC19402;
RA Hofmann F., Habermann E., von Eichel-Streiber C.;
RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z48636; CAA88565.1; -.
DR EMBL; Z23280; CAA80818.1; -.
DR EMBL; Z23281; CAA80819.1; -.
DR PIR; S55805; S55805.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR007577; Gly_transf_sug.
DR Pfam; PF04473; CW binding_1; 10_
DR Pfam; PF04488; Gly_transf_sug; 1.
SQ SEQUENCE 2178 AA; 250134 MW; 9B0ADCE031C4A75A CRC64;

Query Match      27.7%; Score 1435.5; DB 2; Length 2178;
Best Local Similarity 32.8%; Pred. No. 6.7e-47;
Matches 343; Conservative 212; Mismatches 435; Indels 55; Gaps 19;

QY 3 LVNKAQLOQMYVVKRFQDEDEVAAILNALBSEYHNMSESSVVEKYLKDKDINLTDNYL 61
DB 2 LITREQLMKIASIPLKKEPEYNLILDALENFNRDIEGTSVKETYSKLSKLNLVDVNYQT 61

QY 62 TVKSGRNKALKKFKYLTWELKNNSLTPVEKNLHFIWIGGOINDTAINYINQWDX 119
DB 62 KYPSSGRNALLENFRDLSYSELRELIKNSRTSTTASKNLSFINTGSPISQSLSEYNNWK 121

QY 120 DVNSDYTVKVPYDSNAPLINTLTKTIVESATNNTLESFRENLPDPDYKFKRMEIYY 179
DB 122 MFNDYINRLYFDKNSLLVNTLTKTAIQSSSKVIEQNSNLDGTGHNKIFYSDRMKLI 181

QY 180 YDKQGHFDYKSOIEENPEFIIDNIKTLSNEYSKDLEALNKYIEESLNKTANNNGND 239
DB 182 YRYKELKMLYENMKQNS---VDDIINFLSNYFKYDICKLNNQKNNNNKMAIGATD 238

QY 240 IRNLEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVLDILPGIQPLFK 299
DB 239 I-NTENILNTKLKSYVQELIQTNLAAASDILRILKYGKGYVCDLDFLPGVNLSLFN 297

QY 300 SINKPDSINTSWEMIKLEAMKYKEYIPGYTSKNFMDLDEVORSFESALSSKSKSEI 359
DB 298 DISKPNGMSNYEAAIFEAIANEKKLMNNYPYKYMGOVPESEIKERILSFVRN-HDINDL 356

QY 360 FLPLDDIKVSPLEV-----KIAFANNVINOALISLKDSYCSDLVINOIKNRYKIINDNL 414
DB 357 ILPLGDIKLSQELLSLKLKATGKTFSNAFIISNNDLSLTNNLSIQLENRYIILNII 416

QY 415 NP--SINEGTD--FNNTMTKFSDKLASISNEDNMFMKIITNYLKVPAPDVSTINLSG 470
DB 417 QEKFKICETYDSYINSVSELVLETPKNLSMDGSSFYQIIGILYSSGFKPEVNSTVPSFG 476

QY 471 PGVYTGAYQDILLMFKN-----STNIHLPEPELRPFPPKTKISOLTEOITSLWSFNO 524
DB 477 PNIYSATCDTVHFIKNTFDMLSNQOEIPEAS--NNLYFSK-----THDEFKSKSWLRS 529

QY 525 ARAKSOFEYKKGYPEGALGEDNLDPAQNTLVLDKDYVSKII--LSSMKTKNKYIHHIV 582
DB 530 NIAEKEFOKLTKTYIGRTLNYEDGLNFKNKRVTTSLLKVIIBEVNSTKIYENDLNML 589
```

```

QY 583 OLOQDKISYEASCNLFSDPYSSILYQKNIEG-SETAVYVYVADAIEKIDKVRIPYQIS 641
Db 590 QIQGDDISYEAANVFGKPNKSL-----IQGVDDFANVFFENGIVQSDNNILSRFN 645
QY 642 NKRNIKLFTIGHGKSEFNTDFANLDVDSLSSEI-----ETILAKADISPKYIEINLL 696
Db 646 DIKKIKLTLIGHGENVFNPKLFGGKTVDLVNTNIIKPKLOHLLEREGVILKKNYKLNIL 705
QY 697 GCMFYSYISABETYPGKLLIKIDRVSELMP-SISQDSIVTSANOYEVIRINESGKEIL 755
Db 706 GCMFETPKVDYINSTTVGKLFNKIS-----RDLPKFGSKQKLSANKVAIRINREGKREVL 762
QY 756 DSHGKWINKEBIIKDISKEYISPNKENIIVKSKYVHLSLTLQOIRNANSSDIDL 815
Db 763 DIFGKWSNTDIIAQISNKYVYVWNEVENTLSARVEQKLNKVAEPKDI-----NSIIQT 817
QY 816 EKKVWLTECEINVAGNIQRIQVEIRIEAKNLTSDSI-----NYIKNEPKLIES:SDSLYD 871
Db 818 TNNQELKQSLVNTYADLTLLYSLELLKEDIPELQNIQIKERIILNEISRLHDFSNILD 877
QY 872 LKHQGLDSDHSPISFEDISKTEGNGFRIRPINKETGNSPIFETKEIFSEYATHISKELSN 931
Db 878 FYQKNISNMIIILFDSIIKEKDYNVKLANKITGETSVIKTYSDSLWNFTNKYKIVDD 937
QY 932 IKDTIFDVNGKLVKWNLDAAHEVNTLSNAFFIOSLIETNTKESLNSLVAMKVQVYA 991
Db 938 IRGIIKVDINGEFIRKADPEIQNPQLNSAMLMQLLIDYEPYBELTNNMTSLUKVQAYA 997
QY 992 QLPSTGLNITDASKVVELVSTALD 1016
Db 998 QIFQLSIGAIQATEIVTIISDALN 1022

RESULT 11
Q9ZGR4 PRELIMINARY; PRT; 3169 AA.
ID Q9ZGR4 AC Q9ZGR4;
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Putative cytotoxin.
GN L7095.
OS Escherichia coli O157:H7.
OC Plasmid pO157.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EDL933;
RX MEDLINE=98391744; PubMed=9722640;
RA Burland V., Shao Y., Perna N.T., Plunkett G., Sofia H.J.,
RA Blattner F.R.;
RT "The complete DNA sequence and analysis of the large virulence plasmid
of Escherichia coli O157:H7."
RL Nucleic Acids Res. 26:4196-4204 (1998).
DR EMBL; AF074613; AAC70163.1;
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR InterPro; IPR007577; Gly.transf.sug.
DR InterPro; IPR006473; Peptidase_C58_yt.
DR Pfam; PF04488; Gly.transf.sug; 1.
DR TIGRFAMs; TIGR01586; ycpT_cys_prot; 1.
KW Plasmid.
SQ SEQUENCE 3169 AA; 361971 MW; C2CDB48A72BD5FD1 CRC64;

Query Match 8.1%; Score 422.5; DB 2; Length 3169;
Best Local Similarity 21.0%; Pred. No. 4.9e-08;
Matches 258; Conservative 197; Mismatches 450; Indels 323; Gaps 54;

QY 14 YVFERI---QEDFVAILNAL---EYVHNSSESVWEKYKLDKIDNNLTNYLNTYKSGR 68
Db 160 YIKIRTRGAEDQTTTITQSLIINELINGVDRNTI--PFQKISLNDIHSYENMQIKNSR 218

```

```

QY 69 N--KALKKEFYLTWVLELKNNS-----LTPVEKNLHFIWIGQ 106
Db 219 KGEILVKQELLSSLLNDKNGKQSDNASKIINLLGIEYQSHKVDIEPTIHAWVAGA 278
QY 107 INDTAIYINQKVDNSDYTVKVFVDSNAF-----LINTLKKT----- 144
Db 279 PPDNTFSYITAFNTYKYDYTYLLWIDPNAFGAAKFSGILKNIAMNYAIRMURRTNPHLAE 338
QY 145 -----IVESATNNTLE--SPRENLDPEFY-----NKFYRKEMEIYYDQKHFI 187
Db 339 ENNEVLKIQNTONETIEFKEETRERKELENRYKSLTSEKKEFNVFFLESIGMQDNVYF 398
QY 188 DYYKSGIEENPEFI-----IDNIK--TYLSNEYSKDLALNKYIEESLUNKITANNGN-- 238
Db 399 TYCISNGISNTDDISRLDFTLVKLSPVQNDFKSTVEKNKRDIDLLKNTISQKGF 458
QY 239 ---DIRNLKPADELDRLYNQELVERWLAASILIRISMLKEDGGVLDVLDLPGIOP 295
Db 459 QLRDINTLESFKKPDQYFFYOQEMLRWYAAASQVRINILKEYGGIYTDITLPAVSD 518
QY 296 DLPKSIKPDSTINTSWEMIKLEAMKYK--EVIQ--YTSKNFMDLDEVQSFESALS 351
Db 519 KVSQIINE--KSDKRFEDLKLRIISELSILIKGEKYSIKH--DGLDETTLNQLNLI-- 575
QY 352 SKGDKSEIFLPLDDIKVSPLEVKIAFANNSTI-----NQALISLSDSYC 395
Db 576 SEIEK-----LTIDDY--FKPEVETKVVDTFKI PKYQKWTENTWIRGNNNFMLTHKSKC 630
QY 396 SDLVINQIKNRYKILNDNLNPSINEGTFTNTWKIFSDKLASISNE-----DNM 444
Db 631 IDFILSGQKQYLEL--QRIRDNISYNNLFVT-----EDLKSINNVAIGGIPAKKYLEHG 684
QY 445 MFMIKITNVLKVGAFDVRSTINLSCP-----471
Db 685 LF-----SEYRQDGTIPYVYSTLNISGDFMIMRQMKYKSLGRIGEVHDKNKLSDVNFL 740
QY 472 GVTGAYODLLMF---KDNSTNIHLLPE-----LRNPEFPKTKISQLT---EOEITS 518
Db 741 GTVASSNKNKSNFWLNPVSVGINDIPTDDESSWAVRNNDINKILPEKINCHVPEKLPTS 800
QY 519 LWSFNQARAKSQPEYKKGFGALGEDDNLDPQNTVLDKDYVSKILSSMKTRNKEYI 578
Db 801 LY-----YEIDSRSPFGW---DNKSIKHVTEINKOLI--KDNLLLTSSNIDV 844
QY 579 HYIVQIQGDKISYEASCNLFSDPYSSILYQKNIEGSETAVYVYVADAIEKIDKVRIPY 638
Db 845 KLLIKL--DRELYAISSKI-----DNPLALRSIRTLQQLANYVTSTNPENTINFIY 896
QY 639 QISNKN-----IKLTFIGHGKSEFNTDFANLDVDS-----LSSEIET 677
Db 897 DFYRKQDDLLSAIKL-----FSRNDADTKIIVWYNSVMEKNVLEFREVISC 942
QY 678 ILNLAKADISPKYTEINLLGCMNFSYSISABETYPGKLLIKIDRVSELMPISQDSI-- 735
Db 943 VLRSKKVD---SVINEN-----KNLSKEDA--GALRDYAKLKKMELFSLMLDDGYKK 990
QY 736 TVSANQYEVIRINEEGREILDHSGKMNKKEESIIKDISKEYISFNP-----K 793
Db 991 IITINAY---IKERDKL-----SGIIVNIENSIISGHESFDIIRSNOHEWGDLSVTEQFK 1042
QY 784 ENKIIVKSKYLHELSTLQOIRNANSSDIDLEKKVWLFECEINVAS-----NIDQIVE 838
Db 1043 KFEFYVKSE--LSSAKSIFDDIKNK--YITDPTKRNVLHOLDSDIKERIAFLDISHAYP 1100
QY 839 GRIBEAKNLTSDSINVIKNEFKLIESI-----SDSLYDLKHQNGL 878
Db 1101 GSLLEKIQLS-----GYVFSINIIAEVLLASYGVSHGVYVPAPSKLLELRHTK 1156
QY 879 DDSHFIFGEDIKTEGNGFRIRFINKETGNSI---FIETEKEIFSEYATHISKELSNKDT 935
Db 1157 SNSEWI--EKITP-----YVVDIISDNVSVLRPPLSEEQKILNDIKLEISKSVS---EQ 1207

```

Qy 936 IFONVNGKLVKVNLDAAHEVNTLNSAFFTQSIEVNTTK-----ESLSNLSV----- 983
 Db 1208 YFMKLTQKSSVIGIKYKVSDFDRVNEFLSLPLNQMLTLPFMYRYFEMLYDIHIGIEN 1267
 Qy 984 -AMKVQVYAOQLFSTGLNTITDASKVEL 1010
 Db 1268 KANREFIYSKFSLSNLDFLINDERVLNL 1295

RESULT 12
 ID 082916 PRELIMINARY; PRT; 3169 AA.
 AC 082916;
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE Toxin B.
 GN TOXB.
 OS Escherichia coli O157:H7.
 OG Plasmid pO157.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Escherichia.
 CX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7;
 RX MEDLINE=96290540; PubMed=9628576;
 RA Makino K., Ishii K., Yasunaga T., Hattori M., Yokoyama K.,
 RA Yutsudo H.C., Kubota Y., Yamaichi Y., Iida T., Yamamoto K., Honda T.,
 RA Han C.G., Ohtsubo E., Kasamatsu M., Hayashi T., Kuhara S.,
 RA Shinagawa H.:
 RT "Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an
 RT enterohemorrhagic Escherichia coli O157:H7 derived from Sakai
 RT outbreak.";
 RL DNA Res. 5:1-9(1998).
 DR EMBL; AB011549; BAA31815.1; -.
 DR PIR; T00296; T00296.
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
 DR InterPro; IPR007577; Gly trans aug.
 DR InterPro; IPR006473; Peptidase C58_yt.
 DR Pfam; PF04488; Gly trans aug; 1.
 DR TIGRFAMs; TIGR01586; yopT_Cys_prot; 1.
 KW Plasmid.
 SQ SEQUENCE 3169 AA; 361943 MW; D9BED726A62DDBFF CRC64;

Query Match 8.1%; Score 419.5; DB 2; Length 3169;
 Best Local Similarity 20.8%; Pred. No. 6.4e-08;
 Matches 256; Conservative 202; Mismatches 446; Indels 325; Gaps 55;

Qy 14 YKFRFRI---QDEYVAILNAL--EYHNMGESVYKYLKLDINNLTNYLNTYKSGR 68
 Db 150 YIKIRTRGAEQTITITQSLINELLNGVDRNTI-PFQKISELNDIHSVENNQIKNSR 218
 Qy 69 -----NKAALKPKF-----YLTMEVLELKNLSLTPVEKNLHFTWIGG 105
 Db 219 KGIEILVKGELLSLNVKNGKQLSDNASKIINLIGIEYQSHKV-DIEPFFHIAVWAG 277
 Qy 106 QINDTAINYQKWDVNSDYTVKYFYDSNAF-----LINTLTKT----- 144
 Db 278 APPONTFTSYITAFNLTKYKDYTYLLWIDPNAFGAAGKFGILKNIAMNTAIMELRARTNPHLA 337
 Qy 145 -----IVESATNNTLE--SFRENLDNPEFDY-----NKFYKRMELIYDKOKHF 186
 Db 338 EEMNEVILKIQNIQNETIEFKETREKLEENRYKSLTSETKEKFNFFLESIGMQDNY 397
 Qy 187 IDYKSKQIEENPEFI-----DNLIK--TYLSNYSKDLKALNKYIESLNKNTANNGN- 238
 Db 398 FTYCISNGISNTDDISRLDFLTNLVNLKLSPEQNDPFKSTVEKNKRDIDLAKNTISQKFGDR 457
 Qy 239 -----DIRNLEKFAEDIVRLNQELVERWNLAASDILRISMLKEDGGVLDVLDILGIG 294
 Db 458 FQLRDINTLESFKPKPDYFFYQEMLLRWNYAAASDQVRINILKYGGIYDTTDLIPAYS 517

Qy 295 PDLFKSINKPSITNTSWEMIKLEAIMKYK--EYIPG--YTSKXFDMLDEEVQSFBSAL 350
 Db 518 DKVQSILINE-KSDOKRFFEDIKLRIIISSESILSIKGEKYSIKH-DGLDETTLNQLNML 575
 Qy 351 SSKSDKSEIFLPLDDIKVSPLEVKIAPANNSVI-----NQALISLSDSY 394
 Db 576 -SEIEK-----LTIDDY-FKPVETKVRDTPKIFKRYOKWIENTWNIRGNMFMTHKSGK 629
 Qy 395 CSDLVINOIKNRYKILANDNLNPSINEGTDFNTTKIFSDKLASISNE-----DN 443
 Db 630 CIDFILSGCKQKQYLEL-QRIDNISYNNLFYTT-----EDLKSNNVAIGGIPAKKYLEH 683
 Qy 444 MMFMIKITNYLKVGFAPDVRSTINLSGP----- 471
 Db 684 GLF-----SEYRQDGTIPYVWSTLNTISGPDMLRMQMKYKSLGRIGEVHIDKNKLSVNF 739
 Qy 472 -GVYTGAYQDLMF---KDNSTNHLEPE-----LNFEPFKTKLSQLT---EQBIT 517
 Db 740 LGVYASSNKDKNSFNWLPVSGVINDITPDDESWAVANNNDINKILFEKINCHVPEKLPT 799
 Qy 518 SLWSFNQARAKSQPEEYKGFEGALGDDNLDFAQNTVLKDYVSKKILSSMKTRNKEY 577
 Db 800 SLY-----YEIDRSRFFQW---DNKSIKHVTETINKDLI--KDNLLLTSSNID 843
 Qy 578 IHVIVLOQDKISYEASCNLFSDPYSSILYQKNIEGSETAYYYVADAEIKEDKVRIP 637
 Db 844 VKLLIKL--DRELYAISKI-----DNPLALRSIRTQLQLANVYTSNTPEPTINTFI 895
 Qy 638 VOISNKRN-----IKLTFIGHGKSEPTDTFANLDVDS-----LSSIE 676
 Db 896 YDFYRKQDQLLSAIIK-----FSRNDADTKIIVWYNSVMEKNVFLREVIS 941
 Qy 677 TILNLAKADISPKYIEINLLGNMFSYISAEYYPGKLLKIKDRVSELMPISQDSI- 735
 Db 942 CVLRSEKVD---SYINEN-----KKQLSKEDA--GALRDYAKLKMKELFMSLDDDGK 989
 Qy 736 -TVSANQYEVRIINEBGREILDHSGKWINKESIIKDISKEYISFNP----- 782
 Db 990 KIITINAY--IKERDKL-----SGIYINENSIISGHESFDIERSNOHEWGDISTVQEP 1041
 Qy 783 KENKIIVKSKYLHELSTLLOIRNNANSDDIDLEKKVMLTECEINVAS-----NIDROI 837
 Db 1042 KKFEEYVASE-LSSAKSIFDDIKNK-YITDPETKENVLYHQDSDIKERIAFLDISHVAY 1099
 Qy 838 EGRIEAKNLTSDSINYKNEFKLIESI-----SDSLYDLKHONG 877
 Db 1100 PGSILKELQLS---GYVFSIDNIIAEYLLASGVSGHYSHGVVYVAPSPDKLLELRHT 1155
 Qy 878 LDDSHFISFEDISKTEGFRIRFINKETGNSI---FIETEXELFSEYATHISKETSNIKD 934
 Db 1156 KNSSEWI--EKITP---YVYDILSDNVSNLVRLPPLSBEQKILNDIKLEISKVS--- 1206
 Qy 935 TIFDNVNGKLVKVNLDAAHEVNTLNSAFFTQSIEVNTTK-----ESLSNLSV--- 983
 Db 1207 QVFMKLTQKSSVIGIKYKVSDFDRVNEFLSLPLNQMLTLPFMYRYFEMLYDIHIGIE 1266
 Qy 984 -AMKVQVYAOQLFSTGLNTITDASKVEL 1010
 Db 1267 KANREFIYSKFSLSNLDFLINDERVLNL 1295

RESULT 13
 ID 081119 PRELIMINARY; PRT; 3223 AA.
 AC 081119;
 DT 01-OCT-2002 (TReMBLrel. 22, Created)
 DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE Bfal-Lymphostatin-like protein.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Escherichia.
 CX NCBI_TaxID=562;

717 717 -----LKIQRVSELMPSISQDSITVSANOYE---VRINEEG-KREILDHS--GKWINKE 765

1021 QRYLEKMKOHLSLQSSQANEFLISTFIYENDFRKIEAVKNXNNSHELFEKIKKE 108

766 ESIINDISKEY-----ISFNPENKIIIVSKYLHELSLLOEIRNNANSDDIDL 815

1081 QNTWODLTSTKEQLIKALKKEISGNTEKD-----SHYDLLDAPFK--KHNEI-----I 1128

816 EKKVMLTECEINVAS-----NIDROIUEG-----RIBEAKNLTSO-----SINY 854

1129 HNKIQRIODEFKEYSRVAIHNDKVFPGOTDRLHYEGYVFSDFINTLSRYTLHGLGITG 1188

855 IKVEPKLESIDSILYDLKHONGDDSHFISFDISKTENGFRIRFNKGTSGSFITE 914

1189 VHTENLLPAPSSLLINKKE-----HYNEDEISAKPLAYDY-IUNKKESIPVEIL 1241

915 KEIFS-----EVATHISKEISNIKTIFDNVNGKLVKV-----948

1242 NKLSPHELLTPVLGVSNPLGVGYSSDNGKITEQVIYVGADGDPNPSIGLIYTYLED 1301

949 --NLDAAEHVNTLASAFPIQSILEWT-----TKESLNLSVAMKVQVYAQLFS-----T 996

1302 LYNHVRREGTANQ-NLRQLENSVSSCPLTEQSINKILLSEAKEPYOSLTHIOHLT 1360

997 GLNTITDASKVELYSTALDET 1018

1361 GLPTIADAT--LSLSVGLPGT 1380

RESULT 15

Q9RM48 PRELIMINARY; PRT; 3223 AA.

AC Q9RM48;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Lymphostatin.

GN LiFA.

OC Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_Taxid=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=32348/69;

RC MEDLINE=20187489; PubMed=10722613;

RA Klapproch J.M.A., Scalesky I.C.A., McNamara B.P., Lai L.C.,

RA Mastroom C., James S.P., Donnenberg M.S.;

RT "A large toxin from pathogenic Escherichia coli strains that inhibits lymphocyte activation."

RL Infect. Immun. 68:2148-2155(2000).

DR ENBL; AUL33705; CAB55629.1; -

DR GO; GO:0004197; Fcysteine-type endopeptidase activity; IEA.

DR GO; GO:0016740; Ftransferase activity; IEA.

DR GO; GO:0008152; Pmetabolism; IEA.

DR GO; GO:000508; Pproteolysis and peptidolysis; IEA.

DR InterPro; IPRO01917; Aminotrans II.

DR InterPro; IPRO07577; Gly transf sug.

DR InterPro; IPRO06473; Peptidase_C58_Yt.

DR InterPro; IPRO00169; SHprot acsite.

DR Pfam; PF04488; Gly transf sug; 1.

DR TIGRfams; TIGR01586; Yopt cys prot; 1.

DR PROSITE; PS00599; AA TRANSFER_CLASS_2; 1.

DR PROSITE; PS00639; THOL PROTEASE_HIS; 1.

SQ SEQUENCE 3223 AA; 365963 MW; 03E45CEDED7938C1 CRC64;

Query Match 7.4%; Score 386; DB 2; Length 3223;

Best Local Similarity 19.8%; Pred. No. 1.2e-06;

Matches 254; Conservative 206; Mismatches 431; Indels 392; Gaps 55

2 NLVNKAQLQMYYVFRIOEDEYVALNALAEYHNMSSESS-----VVEKYK-LKDINN 54

226 NFVKISLSDELTKY---ANBIIKIKMGYENLLPDKNRNGLKLLQKQADLLKIIME 282

55 LTDNYLVNTYKSGRKALKKPEYLTMEVLEKN---NSLTPVEKNLHFHWGQINDTA 111

283 DTSVTENTPN-----IEMAITDIKEYYSHTVDIEKNTHAIVAGSPESI 329

112 INVINCOWDVNDYTVKFYDSNAFLINTLKTIIVESATNTNLTESFRENLDNPFYKNKF 171

330 SDIYKTFLYTEFTYLLWDKAFGAARFTSVLQAIAFDLACTIQONTFPQKNIDFINL 389

172 Y---RXR-----MEIYDKQHFDY-- 189

390 YNEIRKYNPNPQQGOEXYLNKLRELATYQKISTLPKHMENSFFLENMUKLDNFENVCI 449

190 YKSQIBENPEFIID---NIITYLNE---YSKOLEALNKYIESL-----NKIT 233

450 VKGVTBEINELRINYLNKVIK--LSDDDTIGNYQKTINDKRVKLLDLQKGFGENRIS 507

234 ANNGDIRNLEKFADEDVRLYNQELVERWNLAASDIILRSMLKEDGGVLDVLDILPGI 293

508 IKOVNSLTSLSKENN---HNYQTEMURWNTPAASDLLRWYLKERHGIGITVTDMMPAY 564

294 QPLFLKSINKPDSITNTSMEMIKL-----EALMKYKEVIPGYTSKNFMDLD-EVQRSFE 347

565 SKQVIFKIMQINGNRFLIEDKLRAISDGVRVY-----VNNQNIDEVANYEI----- 613

348 SALSSKSDKSEIFPLPDDIKVPLEVKIAFANSV-----INQ 385

614 ---SDADKNIKKILTISKMPEDSIFTKINTRIPDTMTPIRRYHLWPQGWIRGLNG 669

386 ALISLSDSYCSDLVI---NQIKRYKILNDLNPNSEGTDNTWKIFSOKLASINED 442

670 FMLSHKGSEVIDAVIAGONQARELARIRDNHSEIY----FKQT-----BELSSLPTD 720

443 NMFMKIKITNLYKVG-----FAPDVRSTINLSGP----- 471

721 KIGGIL-VKKYLGSGLSFKPRQDTIIPALSTLQISGPDLIQRMKLOFRGRGVLFGEFI 779

472 ----GVY-----TGAYQ-----DLLMFKNDS 488

780 NERKLSDAKIVGYKITGKYDWLTPESIGYNDVTPADESTWCIGKRGCVDDDFLKDVS 839

489 TNHILLEPELURNPEFPKTKISQLTQEETISLMSFNQAR-AKQPEYKYGHPFGALGDD 547

840 T-----LKTNLPELFLTUKTIDTTFFSQNSTKTKDLQKIQDLTVRYNE--LIDS 889

548 NLDPQANTVLDK-----DYVSKTILLSNMKTENKEYIHHVQLQGDKISYEASNCL 597

890 TIDPKNYLEDOMLHMIMLEMNDIDAKRSLFSLOVQIAEKIRMT-----IPVONIINI 943

598 FSKDPYSSILYQKNIQESSETAYVVYVADAEIKEIDKYRIPYQISNKRNIKULTFIGHKXE 657

944 YPD-----LHKQNDNDLSNISIGFLAS---NPHTKINILYGNKTHEINI----- 983

658 FNTDTFANLDVDSLSSBIETILNAXADISPXYIEINLLGCNNFYSYSIAETYPYKULL- 716

984 FIKOLFPS---FAVMENELRDIIINMSKDKTP-----ENWEGRVNL 1020

InterPro; IPR007577; Gly transf sug.

InterPro; IPR006473; Peptidase_C58_Yt.

InterPro; IPR000169; SHprot acsite.

Pfam; PF04488; Gly transf sug; 1.

TIGRfams; TIGR01586; Yopt cys prot; 1.

PROSITE; PS00599; AA TRANSFER_CLASS_2; 1.

PROSITE; PS00639; THOL PROTEASE_HIS; 1.

SQ SEQUENCE 3223 AA; 365949 MW; CAFD59F69242D07A CRC64;

Query Match 7.5%; Score 389; DB 2; Length 3223;

Best Local Similarity 19.8%; Pred. No. 9.5e-07;

Matches 254; Conservative 205; Mismatches 431; Indels 392; Gaps 55;

2 NLVNKAQLQMYYVFRIOEDEYVALNALAEYHNMSSESS-----VVEKYK-LKDINN 54

226 NFVKISLSDELTKY---ANBIIKIKMGYENLLPDKNRNGLKLLQKQADLLKIIME 282

55 LTDNYLVNTYKSGRKALKKPEYLTMEVLEKN---NSLTPVEKNLHFHWGQINDTA 111

283 DTSVTENTPN-----IEMAITDIKEYYSHTVDIEKNTHAIVAGSPESI 329

112 INVINCOWDVNDYTVKFYDSNAFLINTLKTIIVESATNTNLTESFRENLDNPFYKNKF 171

330 SDIYKTFLYTEFTYLLWDKAFGAARFTSVLQAIAFDLACTIQONTFPQKNIDFINL 389

172 Y---RXR-----MEIYDKQHFDY-- 189

390 YNEIRKYNPNPQQGOEXYLNKLRELATYQKISTLPKHMENSFFLENMUKLDNFENVCI 449

190 YKSQIBENPEFIID---NIITYLNE---YSKOLEALNKYIESL-----NKIT 233

450 VKGVTBEINELRINYLNKVIK--LSDDDTIGNYQKTINDKRVKLLDLQKGFGENRIS 507

234 ANNGDIRNLEKFADEDVRLYNQELVERWNLAASDIILRSMLKEDGGVLDVLDILPGI 293

508 IKOVNSLTSLSKENN---HNYQTEMURWNTPAASDLLRWYLKERHGIGITVTDMMPAY 564

294 QPLFLKSINKPDSITNTSMEMIKL-----EALMKYKEVIPGYTSKNFMDLD-EVQRSFE 347

565 SKQVIFKIMQINGNRFLIEDKLRAISDGVRVY-----VNNQNIDEVANYEI----- 613

348 SALSSKSDKSEIFPLPDDIKVPLEVKIAFANSV-----INQ 385

614 ---SDADKNIKKILTISKMPEDSIFTKINTRIPDTMTPIRRYHLWPQGWIRGLNG 669

386 ALISLSDSYCSDLVI---NQIKRYKILNDLNPNSEGTDNTWKIFSOKLASINED 442

670 FMLSHKGSEVIDAVIAGONQARELARIRDNHSEIY----FKQT-----BELSSLPTD 720

443 NMFMKIKITNLYKVG-----FAPDVRSTINLSGP----- 471

721 KIGGIL-VKKYLGSGLSFKPRQDTIIPALSTLQISGPDLIQRMKLOFRGRGVLFGEFI 779

472 ----GVY-----TGAYQ-----DLLMFKNDS 488

780 NERKLSDAKIVGYKITGKYDWLTPESIGYNDVTPADESTWCIGKRGCVDDDFLKDVS 839

489 TNHILLEPELURNPEFPKTKISQLTQEETISLMSFNQAR-AKQPEYKYGHPFGALGDD 547

840 T-----LKTNLPELFLTUKTIDTTFFSQNSTKTKDLQKIQDLTVRYNE--LIDS 889

548 NLDPQANTVLDK-----DYVSKTILLSNMKTENKEYIHHVQLQGDKISYEASNCL 597

890 TIDPKNYLEDOMLHMIMLEMNDIDAKRSLFSLOVQIAEKIRMT-----IPVONIINI 943

598 FSKDPYSSILYQKNIQESSETAYVVYVADAEIKEIDKYRIPYQISNKRNIKULTFIGHKXE 657

944 YPD-----LHKQNDNDLSNISIGFLAS---NPHTKINILYGNKTHEINI----- 983

658 FNTDTFANLDVDSLSSBIETILNAXADISPXYIEINLLGCNNFYSYSIAETYPYKULL- 716

984 FIKOLFPS---FAVMENELRDIIINMSKDKTP-----ENWEGRVNL 1020

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 16:34:35 ; Search time 10.5 Seconds

(without alignments)
5058.244 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1020

Perfect score: 5189

Sequence: 1 MNLVKAQLQKMYVFRIQ.....ITDASKVVLSTALDETID 1020

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4101	79.0	2366	1	TOXB_CLODI
2	2791	53.8	2710	1	TOXA_CLODI
3	337.5	6.5	2867	1	RBP2_PLAVB
4	282.5	5.4	2869	1	RBP1_PLAVB
5	257.5	5.0	1162	1	EXEN_CLOBO
6	255	4.9	1875	1	MLP1_YEAST
7	251	4.8	2136	1	YCF2_MARPO
8	249	4.8	1679	1	YIO9_YEAST
9	245	4.7	1630	1	MSPI_PLAFK
10	245	4.7	1639	1	MSPI_PLAFW
11	243	4.7	1928	1	MYSL_YEAST
12	242	4.7	1005	1	RA50_METJA
13	241	4.6	1682	1	MSPI_PLAF3
14	240	4.6	1726	1	MSPI_PLAF6
15	240	4.6	1726	1	MSPI_PLAF7
16	239.5	4.6	1957	1	SPOF_SCHPO
17	239.5	4.6	3433	1	UPEO_HUMAN
18	239	4.6	1701	1	MSPI_PLAF5
19	239	4.6	3210	1	CSNF_HUMAN
20	238	4.6	1701	1	MSPI_PLAFM
21	237.5	4.6	1162	1	BYEN_CLOBU
22	237	4.6	1163	1	SPCC_CLOAB
23	236.5	4.6	1730	1	USOL_YEAST
24	236.5	4.6	2399	1	RFC1_PLAFA
25	234	4.5	1179	1	EX5B_BUCAP
26	233	4.5	1225	1	Y309_MYCGE
27	231	4.5	1169	1	SMC_METJA
28	231	4.5	1658	1	YMG7_YEAST
29	230.5	4.4	1727	1	ALM1_SCHPO
30	229.5	4.4	950	1	Y511_RICPR
31	229	4.4	1324	1	SMC4_SCHPO
32	228.5	4.4	1208	1	PCP1_SCHPO
33	228.5	4.4	1290	1	RA50_SCHPO

34 226.5 4.4 1196 1 BXCN_CLOBO P46081 clostridium
35 226 4.4 1175 1 YF19_METJA Q58914 methanococc
36 224 4.3 1803 1 YUL3_YEAST P47024 saccharomyc
37 223 4.3 1102 1 RPOP_AGRBT P33539 agaricus bi
38 223 4.3 1956 1 ATX1_PLAFA Q4956 plasmodium
39 222 4.3 2059 1 TEGU_HSV7J P52362 human herpe
40 221 4.3 1169 1 EX5B_BORBU O51578 borrelia bu
41 220.5 4.2 2104 1 MY93_SCHPO O14157 schizosacch
42 220 4.2 1018 1 YC14_METJA Q58611 methanococc
43 219.5 4.2 1228 1 YHV5_YEAST P38851 saccharomyc
44 219.5 4.2 1805 1 HWM2_MYCGE P47460 mycoplasma
45 219.5 4.2 2663 1 CENE_HUMAN Q02224 homo sapien

ALIGNMENTS

RESULT 1
TOXB_CLODI
ID TOXB_CLODI STANDARD; PRT; 2366 AA.
AC P18177;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Toxin B.
GN TOXB OR TCDB.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RX MEDLINE=90326540; PubMed=2374729;
RA Barroso L.A., Wang S.Z., Phelps C.J., Johnson J.L., Wilkins T.D.;
RT "Nucleotide sequence of Clostridium difficile toxin B gene."
RL Nucleic Acids Res. 18:4004-4004(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RA von Eichel-Streiber C.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1271-2366 FROM N.A.
RC STRAIN=VPI 10463;
RX MEDLINE=92293124; PubMed=1603068;
RA Eichel-Streiber C., Laufenberg-Feldmann R., Sattngen S., Schulze J.,
Sauerborn M.;
RT "Comparative sequence analysis of the Clostridium difficile toxins A
and B."
RL Mol. Gen. Genet. 233:260-268(1992).
CC -!- FUNCTION: Cytotoxin.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; X531138; CAA37298.1; -;
CC EMBL; X92982; CAA63562.1; -;
CC EMBL; X60984; CAA43299.1; -;
CC PIR; A27636; A27636.
CC PIR; S10317; S10317.
CC InterPro; IPR002479; CW_binding.
CC InterPro; IPR007577; Gly_transf_sug.
CC Pfam; PF01473; CW_binding_1; 17.
CC Pfam; PF04468; Gly_transf_sug; 1.
CC Toxin.
SQ SEQUENCE 2366 AA; 269709 MW; E1024BD8B8A56ADF CRC64;

```

RESULT 2
TOXA_CLODI STANDARD; PRT; 2710 AA.
ID TOXA_CLODI
AC P16154;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-PBB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Toxin A.
GN TOXA OR TCDA.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RX MEDLINE=90129305; PubMed=2109310;
RA Sauerborn M., von Eichel-Streiber C.;
RT "Nucleotide sequence of Clostridium difficile toxin A.";
RL Nucleic Acids Res. 18:1629-1630(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RX MEDLINE=90129305; PubMed=2105276;
RA Dove C.H., Wang S.Z., Price S.B., Phelps C.J., Lyerly D.M.,
RT Wilkins T.W., Johnson J.L.;
RL "Molecular characterization of the Clostridium difficile toxin A
gene.";
RN [3]
RP Infect. Immun. 58:480-488(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RX STRAIN=VPI 10463;
RA von Eichel-Streiber C.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION. Only after the enteral delivery of the enterotoxin A may
CC the characteristic disease called pseudomembranous colitis be
CC induced.
CC -!- DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA
CC REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE
CC DIFFERENT OLIGOPEPTIDES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
```

Db 120 INAEYNIKLWYDGEAFVNTLKKAIVESSTTEALQLLEBIEIQNPQPDNMFYKRMFEIY 179
Qy 181 DKQKHFIDYKSOIENPEFIIDNIITKYSNYSKDEALNKYIEESLNKNTANNNDI 240
Db 180 DRQKRFNYIKSQINKEPTVTIDIIKSHLVSYNRDETLESYRNSLRKINSRNGDI 239
Qy 241 RNLEKFADEDLVRLNQLVERWNLAASDILRISMLKEDGGVYLDVDPGLQPLPKS 300
Db 240 RANSLFTEQELNYSQELNLRNGLAASDIVLLAKNFGVYLDVDMPLGHSDFLT 299
Qy 301 INKPDSTINTSWEMIKLEAIMKYKEVTPGYTSKNFDMLEDEQVRSFESALSSKSKSEIF 360
Db 300 ISRPSSIGLDRWEMIKLEAIMKYKYNNTSFNFKLDQQLKDNFKLIESKSEKSEIF 359
Qy 361 LPDDIKVSPLEKVIAPANNVINQALISLSDYSCSDLVNQINQIKNRYKILNINLPSINE 420
Db 360 SKLENVSDLEIKAFALGSVINQALISKQGSYLVNLTVEQVKRYQFNLQHLNPAIRS 419
Qy 421 GTDENTMTKIFSDKLASINEDNMFMKITNLYKGFAPDVSTINLSGPGVYTCAYOD 480
Db 420 DNNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLQGVFPEARSTSLSGPGAYASAYD 479
Qy 481 LLAFKDNSTWHLLEPELRPEPKTKISQTEQITSLWSFNOARAKSOFERYKKGYPE 540
Db 480 PINQENITIEKTKASDLIEFKPENNLSQLTEQINSLSFQASAKYQFEKYVDRDYG 539
Qy 541 GALGEDNLDFAQNTVLDDKY-VSKTILSS--MKTRNKYIHYIVLQGDGKISYEASCNL 597
Db 540 GLSLEDNGVDNKNATLDKNYLLNKKIPSNVBEAGSKNVVHIIQLQGGDISYEATCNL 599
Qy 598 FSKDPYSILYQKNIEGSETAYYYVAD--AEIKEDIKYRIPQIKNKNIKLTFIGHGK 655
Db 600 FSKNPKNSIITQRNV--NESAKSYFLSDGDESILELNKYRIPERLKNKKEKVKVTFIGHGK 657
Qy 656 SEFNTDFANLDVDSLSEIETILNLAKADISPKYIENLLGCMNMSYSISABETYPGKL 715
Db 658 DEFNTSEFARLSVDSLSEISSFLDTTKLDISPKNVEVLLGCMNMSYDFNVEETYPGKL 717
Qy 716 LLKTKDVSSELMPSISODSITVSANOYEVINEEGKEIILDSGKWKINKESIIKDISK 775
Db 718 LLSWMDKITSLPVRKNNSITIGANOYEVINEEGKEIILDSGKWKINKESIIKDISK 777
Qy 776 EYISFNPKENKIIIVKSKYLHELSTLQEIIRNANSDDILEKKVMULTECVINVASNDRQ 835
Db 778 EYIFFSDIDNKLKAKSKNIPGLASISIDITKLLDASVSPDTKFLNLLKLNIESISIDY 837
Qy 836 IVEGRIBEAKNLSDSINYINKNEPKLTESDSLSYDLKHONGLDSDHIFISFEDISKTENG 895
Db 838 IYVEKLEPVKNIHNSIDDLIDFENLVNVSDELYELKLNLDKYLISFEDISKNST 897
Qy 896 FRIRPINKETGNSIFITEKEIFSEYATHISKEISNICKDIFDNVNGKLVKYNLDAHE 955
Db 898 YSVRFINKSGSEVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGLNLDNIQLDHTSQ 957
Qy 956 VNTLNSAFFIOSLIEYNTKESLNSLVAMQVYVYQOLFSTGLNTITDASKVVELVSTAL 1015
Db 958 VNTLNAAFFIOSLIDYSNKKVDNLDSTSVKQVLYAQOLFSTGLNTIYDSQLVNLISNAV 1017
Qy 1016 DETID 1020
Db 1018 NDTIN 1022

RESULT 3
ID_RBP2 PLAVB STANDARD; PRT; 2867 AA.
AC Q00759; Q3N2M3;
DT 01-APR-1993 (Rel. 25, Created)
DT 15-MAR-2004 (Rel. 43, last sequence update)
DT 15-MAR-2004 (Rel. 43, last annotation update)
DE Reticulocyte binding protein 2 precursor (PVRBP-2).
GN RBP-2 OR RBP2.
OS Plasmodium vivax (strain Belem).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A., AND REVISIONS TO 2438-2439.
RX MEDLINE=20299192; PubMed=10838229;
RA Galinski M.R., Xu M., Barnwell J.W.;
RT "Plasmodium vivax reticulocyte binding protein-2 (PVRBP-2) shares structural features with PVRBP-1 and the Plasmodium yoelii 235 kDa rhoptry protein family."; Mol. Biochem. Parasitol. 108:257-262(2000).
RL Mol. Biochem. Parasitol. 108:257-262(2000).
RN [2]
RP SEQUENCE OF 1189-2439 FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax merozoites."; Cell 69:1213-1226(1992).
RL Cell 69:1213-1226(1992).
CC -!- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to human reticulocyte cells.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.ebi.ac.uk/ebis/announce/ or send an email to license@ebi.ac.uk).
CC EMBL: AF184623; AAF76525.1; -
DR HSP; P03069; IGM; Transmembrane; Repeat.
KW Malaria; Receptor; Signal; POTENTIAL.
FT SIGNAL 1 21
FT CHAIN 22 2867 RETICULOCYTE BINDING PROTEIN 2.
FT DOMAIN 22 2805 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2806 2826 POTENTIAL.
FT DOMAIN 2827 2867 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 44 133 ASN-RICH.
FT DOMAIN 560 758 LYS-RICH.
FT DOMAIN 1112 1285 LYS-RICH.
FT DOMAIN 2758 2785 7 X 4 AA TANDEM REPEATS OF H-D-D-T.
FT REPEAT 2758 2761 1.
FT REPEAT 2762 2765 2.
FT REPEAT 2766 2769 3.
FT REPEAT 2770 2773 4.
FT REPEAT 2774 2777 5.
FT REPEAT 2778 2781 6.
FT REPEAT 2782 2785 7.
SQ SEQUENCE 2867 AA; 331433 MW; 6E7D8CA71AFBFFD3 CRC64;
Query Match 6.5%; Score 337.5; DB 1; Length 2867;
Best Local Similarity 19.8%; Pred No. 1e-05;
Matches 250; Conservative 226; Mismatches 454; Indels 335; Gaps 57;
Qy 2 NLV-NKAQKQKVVYKFR-----IQDEYVAILNALALEEYHNMS----- 40
Db 977 NIIRKADLIKKLDQHTQIEKHFTFIEEMSPLLSVIKKCKRVESDWSELIKQLNTK 1036
Qy 41 --SVVEKYLKLD-----INLTDNYLNTYKSGENKAL 72
Db 1037 INAILLEYNKSCKDRFNGDDTNLELDDFKKQDAQOEIKKLTNY--NVLNDGINVII 1094
Qy 73 KKFKEYLTWVLELKNNSITPVKKNLHFIWIGQINDTA---INVINQWK-----DVNS 123
Db 1095 KEQHE---KVIILSENHITKDK-----KINEKIQQNVNLSNEMKTKLGLLKINE 1141
Qy 124 DVTVKVYDVSNAFLINTLKTIVE-----SATNNTLESFRE--NLNDPEFDYN 169
Db 1142 D--IKNSRDT-----TKSKIQEKKVQVTFGSIDYANKKIDAKKEHDVNDKEDFKE 1193
Qy 170 KF-----YRKRMEIIVDKQKHFII-DYKQSIQIENPEFIIINIIKTYSNYSKDLAL 221
Db 1194 KYKDTSFDEKKKSIKAYEKMGNTLKELEKMDDEKNIEKEVEAQIQYKRFIDHVDNLM 1253

```
QY 222 NKVIESLNKITANNND:RNLEKFADEDLVRLYNQEL-----VE 261
Db 1254 NDEVEKS--KIV-----MEK-----IELYKEIDEIKQKNEVKQDGTNFYITE 1296
QY 262 RWNLAAS-----DILRISMLKE-DGVVYLDVILPGIQPOLFSINKPDSITNTSWEM- 314
Db 1297 QYNSATQSKAKIEQFINIATTKGTSDTSODINELLESKEEVHKQLVLVKQESNSMEBR 1356
QY 315 -----IKLEIMKYKEYIPGYTSKN-----FMDLDEVORSFESALSSKSD 355
Db 1357 KQILSKMDLLINNSSETIAKESNTQNALGFRENAKTKLNKTDELLQKVAAMTEEAKAH 1416
QY 356 KSEIFLPDDIKVSPLEVKIARNSVINQA-----LISLKD--SYCSDIVINQIKRY 407
Db 1417 KNNIDIALEDAQIDTEVSKIEQINREIMNMKDEIKSYLSEIKYKDKCTTISNSKRGKD 1476
QY 408 KI-----LNDNLNPSINEGTDF-----NTWKIFSDKLASIS 439
Db 1477 KIEFLKFPNBSNSKNVINEINENINSQYIKDIEDARKQASTKVELFPHKHTTIS 1536
QY 440 N-EDNMFMKITYNLKVGFPADVRSTINLSGPGVYT--GAYQDLLMFKONSTNIHLE 495
Db 1537 NIFKESEILGVETKSKINKAEDIMKEIERHENSEIQTKVGFOENKLNKNEPHYDNAE 1596
QY 496 PELRNFEPTKISQTEIITSLMSFNQARAKSQFEYKGYFEGALGE---DNDLQFA 552
Db 1597 DELNNDKSTNAKV--LIETNLSV-----KHNLSBITNIKQ-----GEKIYSKADIM 1643
QY 553 QNTVLDKDYVSKILSSMTRNKYEIHYIVQLQGDKISYEASCNLPSKDPYSSI--LYQKN 611
Db 1644 QKIKATSENTAEKTEKVKDDQSNVNYL-----NQITERNLIVTEKNRLNGIDSTITN 1698
QY 612 IEQS--ETAYVYVADAE--IKEIDKIRIPQISNKRNIKLTIFHGKSEFNT-----DT 662
Db 1699 IEGALKESKNGVEIGLEKLEIGKRLKVDITKXSINST-VGNFSSLFNNFDLNOYDF 1757
QY 663 FANLD-----VPSLSSEIETILNLAADI-----SPKYI-----EINLLG-- 697
Db 1758 NKQINDYENKMGIEYNEFGSLNKISENLARNASENTSDYNSAKTLRLAEQKKNLLNKE 1817
QY 698 --CNMFYSISABEETPGKLLIKIDRVSELMPSISQDSITVSANQYEVRIINEBKREIL 755
Db 1818 EEANKYLRDVKVKESEF--RFIFNMKESLDKINEMIKKEQLTYNEGHGNNVKQJLVENIKELV 1875
QY 756 DHSGKWKNEESTIKDISKEVYSFNPKENKILIVSKYLHELSTLLOERNNANSDDLD 815
Db 1876 DE-----NLSLDLKQATGN-----BEIQKITHSTLKNKAKTILGHVDTSKTVGKI 1924
QY 816 EKKVWLTE-----CEINVASNIDRQIVEGRIEBAKNLTSDSINYIKNEFKLIBESID- 867
Db 1925 TPELALTELLGDAKLTAQELKFEKSNVVLTEENMSKT-----NELDVHKNIQDAYKV 1979
QY 868 SLVDLKHQGLDSDHIFSEDISK-TENG-----FRIRFINKETGNSFIETEKETFSYA 922
Db 1980 ALEILAHSDIDTPQ---KDSKLEMGNQIVLVKVLINQYKNKISSIKSKEEAVSVKI 2035
QY 923 THLSK--EISNI--KDTIFDNGVKLVKKNLDAHEVNTLNSAFFIQ-----SLI 969
Db 2036 GNVSKHSELSKITSCKSYDNT-----IALEKQTELQNLNSFTQEKTNNSDSKLE 2088
QY 970 EYNTKESISNLSVAMK-----VOVYAOQFSTGLNTITDASKVBLVSTA 1014
Db 2089 KIKTDFESLKNALKTEGEVNALKASDNDHEHVQSKSEPNPALSIEKEETDIDSLNTA 2148
QY 1015 LDET1 1019
Db 2149 LDELL 2153
```

RESULT 4
RBPI_PLAYB
ID_RBPI_PLAYB

```
AC Q00798;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Reticulocyte binding protein 1 precursor.  
GN RBPI.  
OS Plasmodium vivax (strain Belém).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI_TaxID=31273;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92315338; PubMed=1617731;  
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;  
RT "A reticulocyte-binding protein complex of Plasmodium vivax  
merozoites";  
RL Cell 69:1213-1226(1992).  
CC -!- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to  
human reticulocyte cells.  
CC -!- SUBUNIT: Homodimer (Potential).  
CC -!- SUBCELLULAR LOCATION: Membrane-bound.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
CC -----  
DE EMBL: M88097; AAA29743.1; -;  
KW Malaria; Receptor; Signal; Transmembrane.  
FT SIGNAL 1 17 POTENTIAL.  
FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.  
FT DOMAIN 18 2807 EXTRACELLULAR.  
FT TRANSMEM 2808 2826 POTENTIAL.  
FT DOMAIN 2827 2869 CYTOPLASMIC.  
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).  
SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205BECFF CRC64;  
  
Query Match 5.4%; Score 282.5; DB 1; Length 2869;  
Best Local Similarity 20.0%; Pred. No. 0.0014;  
Matches 242; Conservative 212; Mismatches 410; Indels 349; Gaps 61;  
  
QY 6 KAQLQKVVYKFIQIDVEY-----VAIINALEEYHNMSESVVVEKYLK-LK-DINNL- 55  
Db 946 KALKEKIVSDSLDKIDQYETEFKEKTSVAVENTVSTIQSLSKA--IDSLKRLNGSINNCK 1003  
QY 56 ---TD-NYLYNTYKSGRNKALK-----KFKBYLNEVLELKNLSLTPVKNLHFIWIG 104  
Db 1004 KYNTDIDLLRSKIKTLREEVQKEMPRKGDCKGENTTALLKSLDRKMGKINEXLN---D 1059  
QY 105 QGIN--DT-----AINVINQWKVN----- 122  
Db 1060 GRNLSDLTCKEDLLKPYFSESKSIHLKSKQKQGPDPANRIDEWEDIKRDVDELNVNQVI 1119  
QY 123 SDYTVKVFYDSNAFLNTLKTTI-----VESATNTLTESFRE-----NLNDPEFYDNKF 171  
Db 1120 SENKVLTFKNSVTYIEAMHSHINTVAHGITSNKLKSVKEVDEKLNLEQVEDYKVV 1179  
QY 172 ----YKRMREII---YDKQKHFDIYKSGIENPEFIIDNIKTYLSN-----EYSKLEA 220  
Db 1180 KNPENEQLEAIRGSMKLEKINKHVSEMTQ-----LESTANTLAKSNAGKENEHDL 1234  
QY 221 LNKY-----IEESLNKI-----TANNGNDRNLKEFADEDLVRLYNQELVERWNLA 267  
Db 1235 LNKTKGQMRDIYEKLLKIABELKEGTIVNELKDNANKVPEPEPFIITGHVLE----- 1288  
QY 268 ASDILRISMLKEDGVYLDVILPGIQPOLFKSINKPDSITNTSWEMIKLEIMKYKEYI 327  
Db 1289 ----RITVEKQKAGV--VEEMNSLTKIEKLIQE---TSDDSQNELVTTSTIKHLENA 1338  
QY 328 PGVTSKNFMDLDBEVQSFESALSSKSDKSEIFLPLDDIKVSPLEVKI----AFANNSVI 383
```



```
Db 728 K-----FTSMEQCKININKSTKEPILKCTNINETEKSHLIMQNSFSLNDFDL 776
Qy 590 SYEASCNLFSDPYSSIIYQKNGESATAYYY-----VADAEIK-----EIDK-YRIP 637
Db 777 DIQNMKNLFNL--YTELLAIKQTSPEYLSLYAFQEQDNNVIGDTSKNTLVVEPKDIGLV 834
Qy 638 YOISN-----KXNIKLT--FIGHGKSEFTDTFA-----NLDVDSLSEI----- 675
Db 835 YGINNAHILGANCQNIKFTNDYFENGJ-----TNNFSIYFWLRNLKQNTKSKLIGSKED 890
Qy 676 -----ETILNLAKADISPKYIBINLLGCMFMSYSISASBETYPGKILLIKIDRVS 724
Db 891 NCGWEIYFENDGLVFNIDSGNEQNIYLSNKNKSWHYIVISINRLKDQLLIFI----- 945
Qy 725 ELMPSIODSITVSANQVEVRINECKREIHDGKWKINKEESIKDISKEYISFNPK 784
Db 946 -----DNILVA-----NEDIK-EIL-----NIYSDIISLSDN 973
Qy 785 NKIIYVSKYLHLSLTLQEIIRNA-----NSSDIDLEKKVMLTECEINVASNIDR 834
Db 974 NNV-----YIEGLSVLNTIINSNEILTYFSDLNNSYIRNFDSEILQVNRTYELNFVFP 1028
Qy 835 QVEGRIBEAKNL-----TSDSINYIKNEFKLIESISDSLYDLKHQ-----NGLDSDSHFIS 885
Db 1029 EIAINKIEQNNIYLSINNENNLNFKPLKLLNTPNKQYVQKWDVIFSVLDGTE--K 1086
Qy 886 FEDISKTENGPIRINKETGNSIFITEKEIFSEYATHISKEISNIKOTIFDNVNGKLV 945
Db 1087 YLDISTTNN--RIQLVDNKNNAQIFI-INNDIF-----ISNCLTYNNVNVVLS 1133
Qy 946 KK-----VNLDAAEV 956
Db 1134 IRNQDYNWVICDLNHI 1150

RESULT 6
MULTI_YEAST
ID MULTI_YEAST STANDARD; PRT; 1875 AA.
AC Q02455;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin-like protein MPlI.
GN MPlI OR YKR095W OR YKR415.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94205265; PubMed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UB12 and MPlI genes and three
RT new open reading frames.";
RL Yeast 9:1349-1354(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94205265; PubMed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UB12 and MPlI genes and three
RT new open reading frames.";
RL Yeast 9:1349-1354(1993).
RN [3]
RP FUNCTION: Myosin-like protein that is probably involved in DNA
RP repair.
CC -! SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC -! CAUTION: Ref.2 misquotes the gene name as "MPlI".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L01992; AAA34783.1; -
DR EMBL; X73541; CAAS1948.1; -
DR EMBL; Z28320; CAAB2174.1; -
DR PIR; S38173; S38173.
DR GenOnline; 140074; -
DR SGD; S0001803; MPlI.
DR GO; GO:0005635; C:nuclear membrane; IDA.
DR GO; GO:0005654; C:nucleoplasm; IDA.
DR GO; GO:0006066; P:protein-nucleus import; IDA.
KW Coiled coil; DNA repair.
FT DOMAIN 69 487 COILED COIL (POTENTIAL).
FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
FT CONFLICT 301 301 R -> A (IN REF. 1).
SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C906867 CRC64;

Query Match 4.9%; Score 255; DB 1; Length 1875;
Best Local Similarity 17.4%; Pred. No. 0.0094;
Matches 234; Conservative 241; Mismatches 432; Indels 436; Gaps 55;

Qy 14 YKFRIQDEYVAIL-----NALEYVHNMSESSVVEKYLKLDINN 55
Db 225 YLSYRKTKVILDIRNELNRLNDFQWERTNVDVKQNNLSKSLQEKLEIK---GL 281
Qy 56 TNYLNTYKSGRNKALKKFKYLTMEVLELKNLSLTPVEKNLHFIIWGGQINDAINY 115
Db 282 SDSLSSEKQEPSAEMSLQ-----RLVDLLESQNLNAKVEL----- 317
Qy 116 NQWQVNSDYTKVFD-----SNAFLINTL-----KKTIVESTATNVLBSF 157
Db 318 NSIRELN--TAKVIADDSKQTPENEDLLKELQTLKEKLAQCEKELSLSSITDEAD 374
Qy 158 RENLNDPEFDYKFKRMEIIVDKQHFIDYKQIENPEFIID-----NIKTY--- 209
Db 375 NENLSAKSSDFILKKQL-----IKERTKEHLQVQIE--TFIVELEHKVPIINSFK 427
Qy 210 ---LSNEYS-----KOLEALNKVIEESLNK----- 231
Db 428 TMLNELNNAALLLEHTSNEKNKAKVELNAKNQKLVCEENDLQTLTKORLDCRQIQL 487
Qy 232 -ITANGND-----IRNL-----EKPADELVLRYNQBELVRWNLAASDILR 273
Db 488 LIITSVNSDKGFLRKEEIQFIONIMQEDDSTITESDSQKVYTERLVEFKN----- 538
Qy 274 ISMLKEDGGVYLDVILPGIQDLPKSNKPSINTSWEMIKLEAIMKYKYEIIPGYTSK 333
Db 539 IIQLEKNAELLKV-----VRNLADKLESKEKSKSLQKIESETVNAEKAIIITLKE 592
Qy 334 NFDMLD--BEVQSPESALSCKSKSEIFLPD-----DIKVSPLSVKI----- 375
Db 593 KMDLESRIEQLKELEELKTSVPNEDASVNTIKQITETKDDLESQVODLQTRISQITR 652
Qy 376 -AFANNSVINQALISLSDYSCDVLNQIK-----NRYKILNDNL-----NPSIN 419
Db 653 ESTENSLNKEIQDLYDSK-SDISIKLGKSSRLAEERFKLSLNTLDTLKAENDQLR 711
Qy 420 EGTDF--NTTMKIFS-----DKLASIGNEDNMFMKITYNLKVGFPADVRS 464
Db 712 KRFYDLQNTILKQDSKTHETLNEYVSCSKLSIVETE--LNLKEEQKLRVHLEKNLQ 768
Qy 465 TYNLSPGVYTGAYQDLMFKONSTNIHLPELRNFPFKTKISOLT-----EQEITSIN 520
Db 769 ELNKLSP-----EKSLRIMVTLQTLQKEREDELEETRSCKKIDEL- 812
Qy 521 SFNQRAKQSFQYKGYFEGALGEDD--NLDFQNTV--LDKDYVSKKLSLSSMKTRNKE 576
Db 813 --EDALSELKQTSQKDHIIKIQLEEDNNNIEWQNKIEALKADYESVITSVDKQTDIE 870
```

```
QY 577 YHIV-----QLOQDKI---SY-----BASCNFSPKDPYSI-----606
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 871 KLOQVKVSLKEIEEDKIRLHYTNVMDTINDSDSLRKELEKSNL--TDAYSOIKKEYKD 928
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 607 -----LYQKNTEGSE-----TAYYYVADAEIKEDKYRI-----PYOISKNRIKLT 649
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 929 LYETTSOQNTSKLDESFPDFTNQIKNLTDEKTSLEDKISLKEQMFNLNELDLQK 988
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 650 FIGHKGSEFNTDTFA-----NLDVDSLSEIETILNLAADISPKYIEINLLGCNMFYSYI 705
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 989 GMEKEKADFKRISILQNNKEVAVKSEYESKLSKIQNDLDQGTI-----YAN 1037
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 706 SAETYPGKLK--LKIKDRVSELMPSI-----SODSITVSANQYVEINBEGREILDHS 758
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1038 TAQNNYEQLQKHADVSKTISELREQHLYKGQVKTNLNLRDQLENALKENEX-----1090
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 759 GKNWINKBESIIDS--KEYISFNPKNKIIKYSKYLHELSTLLOEIRNANSSD-----812
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1091 -SWSQSLEQLDLSNRISLSSQNKLLYDQIQIYTAAD--KEYVNSTNGPGLNII 1147
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 813 IDLEKMYLTCENVA--SNIDRQ--IVEGRIEBAKNLTSDSINVIKNEFKLIESIS 866
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1148 ITLRERDILDTKVVAERDAKMLRQKLSMDVQLQARTKLDNSRVEKENHSSIIQOHD 1207
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 867 DSLVDLKHQGLDSDHFTSPEDISKTENGFR-----IRFINKE 904
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1208 DIMKQLNLLRESNITRLNELNENNNKKELQSELDKLQNVAPIESLTLAKYSMQE 1267
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 905 TGNIS-----FITEKEIFSEYATHISKEISNIKOTI-----936
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1268 KEQELKLAKEVHRWKRQSDILEKHEQLSSSDVEKLESELENKELENKERQGAEE 1327
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 937 -FDNVNGKL-----VKKNLDA-AHEVNTLSAFTI--QSLIENTYTKESLSNLVAM--985
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1328 KFNLRQAQERLKTSLQSDSLTEQVNSLRDAKNVLNLSSEANARIELQNAKVAQGN 1387
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 986 -----KQVYVAQLF 994
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1388 NQLEAIRKLOEDAKASRELOAKULEBSTTSVESTINGINEITLKEIEKORIQOOLQ 1447
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 995 STGLNTITDASKWVELVSTALDE 1017
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1448 ATSEANEQNDLNVESMKSPFE 1470
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
ID YCF2 MARPO STANDARD; PRT; 2136 AA.
AC P09975;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical 259.9 kDa protein ycf2 (ORF 2136).
GN YCF2.
OS Marchantia polymorpha (Liverwort).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Eukaryota; Viridiplantae; Marchantiidae; Marchantiales; Marchantiaceae;
OC Marchantiaceae; Marchantia.
OX NCBI_TaxID=3197;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohmawa K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
RA Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,
RA Ozeki H.,
RT "Chloroplast gene organization deduced from complete sequence of
RT Liverwort Marchantia polymorpha chloroplast DNA."
RL Nature 322:572-574 (1986).
CC -!- SIMILARITY: Belongs to the ycf2 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; X04465; CA28078.1; -.
CC PIR; S01591; A05037.
CC InterPro; IPR003593; AAA_AtPase.
CC InterPro; IPR003959; AAA_AtPase_cent.
CC Pfam; PF00004; AAA; 1.
CC SMART; SM00382; AAA; 1.
CC Chloroplast; Hypothetical protein.
SQ SEQUENCE 2136 AA; 259911 MW; 5BD170C9CCF61197 CRC64;
Query Match 4.8%; Score 251; DB 1; Length 2136;
Best Local Similarity 21.0%; Pred No. 0.016;
Matches 227; Conservative 153; Mismatches 329; Indels 374; Gaps 55;
QY 27 ILNLEEEYHNSSESVVEKYLKLDINNL-TDNYL-----NTYKSGRNKALKKPKYLT 80
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 309 IFKNLQNF-NESDKLIESFFLLKIKGNLYFQNYIEFTWQSYKDKCLD--FNKFNELNN 365
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 81 MEV-LELKNNSLTPVEKHLFIWGGQINDTAI-----NYINQWKDVNS--DYTVKVFVDS 133
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 366 SEIYIKIELFSDYTYKFSKYLVBGKSKTIIKQSFNNNIYKYLNSIFNFTIFYFDS 425
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 134 NAFNLTKKTIVESAATNT--LESFRENLDPEFDY---NKFYRKEMELIYDKQHEFI 187
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 426 NLLFDWLKQNY--INNKPLKSF-----LIYSSNQF-----ILFFKQKSK 468
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 188 DYYKQIENPEFIIDNIKTLS--NEYSKOL-----EALNKVIEESLNKIT 233
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 469 SFNKLKVKQSDVITVNFSEKNEKIEINNFSKIYVAFELISINEIDNKFV--INKIS 525
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 234 ANNGDIRN---LEKFADEDLVRLYNOELVERWN-----LAAASDILRIS-ML 277
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 526 LKNINKKQKRYLKNKIKSSDNFRFLNWKIKGVSSQOFVSNNSFLLNPAPEILOQNYL 585
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 278 KEDGGYLDVDLPGIQDPLFKSINKPDS-----TNTSWEMIKLEAIWK----322
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 586 KKKNLIF-----FKKLEVFNSFFYQYKCKKLNIFLKFALEKILKQNK 632
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 323 -----YKEYIPGVTSKNFMDLDEEVQSRPESALSSKSDKSEIFLPLDDIKV-----368
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 633 KFTISIKLFFKFKYKNKLNENGEYKIESQILQNEKELNKKRKNQFQNP--NIKISFYNS 690
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 369 SPLEVKIA---FANNVINQALISLKDYSCLDVINQIKRYKILNDNLNPSINEGTFN 425
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 691 SKKNYILQNKYFFKNLNNKILTWK-----KISNKLVISNSEN-----730
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 426 ITMKIFSDKLASISNEDNMFMKITNYLKVGFADPVRSTINLSGPGVYTGAYQDLLMPK 485
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 731 ---KI-----INWKNMKP-----FSPSKNSVL-----DTFFFN 756
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 486 DNSTNIHLEPELRNFBEPKTKISQITQBEITSLSW-----PNQARAKSQPEEYKGYFE 540
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 757 KKSFNITV-----IFDLKAKIQNFQEIQKILNCFSLFNSKNIK-KTKIFKNSYF- 807
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 541 GALGEDNDLFAQNTVLDKDYVSKILSMKTRKEYIHYIQLQGDKISYEASCNLPSK 600
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 808 -----INENLTTFSFNDKEFNIFLFISELNDFLNRFFAK 846
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 601 DPYSILYQKNEGSETAYVYVADAET--KEIDKYRIPYQISNKRNIKLTPIGHGKSEF 658
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 847 -----YLYRYIKKKEILFNPIENRQLLQNFETKTI-LTFI-----DF 884
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 659 NDTTFANLDVDSLSSEIETILNLAADISPKYIEINLLGCNMFYSISABETYPGKLK 718
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 885 LQD-----PELNNYNNRIFLPHLEKTIKQV--NLL-----YLRLLKIF 919
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 719 IKDRVSELMPSISQDSITVSANQYVEINBEGREILDHSGKWNKERSIIDKISSKEYI 778
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 920 LKDKRNFLL-----INE-----IK-----933
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

QY 779 SFNPKENKIIIVSK-----YLHELSTLLOEIRNANSSDIDLEK 817
Db 934 SFIEKKNLFIKQSLSNVLNKNYSKFFDNIFHLKQKEKNIEIILNKNQYFEKSLK 993
QY 818 KVMLTECEINVA-SNIDROIVEGRIEAEKNLTSDSINTVKNKFLIESDSLSY-----D 871
Db 994 KYLKNLNLNNSYSKFSYKIF---IFOLLNLN-----KNYKTFQWISLFFYSKNLN 1044
QY 872 LKHQGLDSDHSPISPEDISKTENGFR-IRFINK-----ETGNSIFETEKEIFSEYATHIS 926
Db 1045 YKIQNKIEKNYCNKYNISYKKKIKTNVFFKKNLFTNNSWFFTL-----WWEYNTYIL 1101
QY 927 KEISNICKDTIFDNV-----GLVKVYNLDAAEVNTLSAFAFIQSLEIENYTKESLS 979
Db 1102 LQI--IQETFFQITVLEVFYKKKIKIEK-----NLKFFLKS-----KKISLK 1141
QY 980 NLS 982
Db 1142 TLS 1144
RESULT 8
YIO9 YEAST STANDARD; PRT; 1679 AA.
AC F40457;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical 195.1 kDa protein in DNA43-UBI1 intergenic region.
GN YIL149C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313266; PubMed=9169870;
RA Churche C.M., Bowman S., Badcock K., Bankier A., Brown D.,
RA Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,
RA Harris D.E., Horsnell T., Hunt S., Jagels K., Jones M., Lye G.,
RA Moulle S., Odell C., Pearson D., Rajadream M.A., Rice P., Rowley N.,
RA Skelton J., Smith V., Walsh S., Whitehead S., Barrell B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX";
RL Nature 387:84-87(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z386129.1; -
CC PIR; S48385; S48385.
CC GernOnline; 139684; -
CC SGD; S0001411; MLP2.
CC GO; GO:0005635; C:nuclear membrane; IDA.
CC GO; GO:0005654; C:nucleoplasm; IDA.
CC GO; GO:0006606; P:protein-nucleus import; IDA.
CC KW Hypothetical protein.
SQ SEQUENCE 1679 AA; 195141 MW; 298950CC52202D8F CRC64;
Query Match 4.8%; Score 249; DB 1; Length 1679;
Best Local Similarity 19.7%; Pred. No. 0.014;
Matches 226; Conservative 193; Mismatches 373; Indels 356; Gaps 61;
QY 3 LVNKAQLQKMYKFRIOQDEYVAILNALEEVH-----NMSSESVVEKYLKLDINNLT-D 57
Db 469 LNTSAIQE---TASPLSQDELISLRKILESNNIVNDSQAIITRLVFEFSNVNELQEK 525

QY 58 --NYLNTYK-----KSGENKALKKPKFYLTME-----VLELKNNSLTPVERNLHFIW 102
Db 526 NVZLLNLCIRLADKLENYEGKQDKTLQKVENOTIKEADAIIELEN----- 571
QY 103 IGOQIN---DTAINYNQWKNVSNVYKVFYDSNAFLINTLUKKTIVESATNNTLESFRE 159
Db 572 ---INAKMETRINILLRERD-----SYKLASTEENKANTNSVTSMEAREKIKIRELEA 622
QY 160 NLNDPFDYKFKRMEIYYDKQHFIDYKYSQTEENPEFIDNIIKTYLSNEYSKDLLE 219
Db 623 ELSSTKVE-----NSAIQNLKELLTYKKSQCKK-----KTTL----- 556
QY 220 ALNKYTEESLUNKITANNGNDIRLEKFADEDLVLYNQBELVERWNLAASDILRISMLE 279
Db 657 -----EDFENFKGLAKEK-----ERM-LEEADHILKAELEKQ 687
QY 280 DGGV--YLDVILPGIQDPLFKSINKPDSITNTSWEMIKLEA---IMKYKEYIPGYTSK 333
Db 688 KSWVPSYIHYE-----KERASTELSQSRIKIKSLEYEISKLKK-----ETA 728
QY 334 NFDMLDEEVORSPESALSCKSKSEIFLPLDDIKVSPLEVKIAPANNVINQALISLKDS 393
Db 729 SFPTKESLTRDFEQCC---KEKELQMLKKESEISHNENKMDFSKSGQYKA----- 778
QY 394 YCSDLVINQIKRYKILNDNLNPSINE-----GTFDNTMTKIFSDKLAS 437
Db 779 -----KIKELNLERLRSLOSKIQIEISIRCKDSQLKWAQNTIDDTMKMKLS-LLTE 832
QY 438 ISNEDNMFMII---KITNLYKVGPAQVRSITNLSPGVVTGAYQDILLMFKNSTNHLLE 495
Db 833 LSNKETTIEKLSSEIENLDK-----ELRKT-----KFKQ--YKFLDQNSDASTLE 874
QY 496 PEIRNEFPFKTKI-----SQL-TEQETISLWSFNQARAKSQPEEYKGY----- 538
Db 875 PTLRK-ELEQIQVLQKANSQIQAYEEIITSSNENALIELKNELAKENYDAXIELEKKE 933
QY 539 -----FEGALGE-----DDNLDPAQNT-----VLDK----- 559
Db 934 KWAREEDLSRLGELGEIRALQPKLKEGALHFVQSEKLRNEVERIQRMKIEKMTIV 993
QY 560 DYVSKTILS---SMKTRNKYIHYIVOLOQDKISYEASCNLFPSKDPYSSILYOKNIEGSE 616
Db 994 QLCCKKMSQYQTMKNKDLSELVIRLEKDAADCAELTKTKSSLYSA---QDLDDKHE 1050
QY 617 TAYYYYVADAE-----IKEIDKIRIPYQI-----SNKRNKILFTFHGKSEF 658
Db 1051 RKWMEERADYERELISNIEQTESLRVENSVLIEKVDDTAANNQKQHLKLVSL----- 1103
QY 659 NDTFANLVDLSSEIETILNLAADISPKYIE-----INLLGNMFSYSISAE 709
Db 1104 ---FSNLRHE---RNSLETMLTKCKELA---FVKQKNSLEKTINDL---QRTQTLSEKE 1152
QY 710 TYPGKLLLLKIDRVSELMSPISQDSITVSAN---QVEVRAINEEGKEILDHSGKWINKEE 766
Db 1153 YQCSAVII---DBFKDITKEVTQVNIILKNNAILQSKLNKVTKEKREIY---KQNDKQ 1205
QY 767 SII-----KD-ISSKEYISFNPKENKIIIVKSKYLHELSTLLOEI---RNNANSSDID-LEK 817
Db 1206 EETSLRQDLQITKEQVSN---SNKILVYSEMEQCKQRYQDLSQOQKQAKQDIEKLN 1263
QY 818 KVMLTECEINVASNIDRQIVEGRIEAEKNLTSDSINYIK-----NEFKLIESDSLS- 869
Db 1264 EISDLKGLKLSAENANADL-ENKFNELKKAQHEKLDASKQQAALTNELNELKAIKDKLE 1322
QY 870 YDLKHQNG-----LD-----DSHFISFEDISK--TENGFR-----IRFINKE-----TQNSIF 910
Db 1323 QDLHFENAKVIDLTKLKAHELOSVDYRDHEKDTYRTLMEELESUKLEQIFKTANS-- 1380
QY 911 IETEKEIFSEYATHISKEISNICKDTIFDNVNGKLVKNLDAAEVNTLSAFAFIQSLE 970
Db 1381 ---SSDAFEKLVNMEKE-----KORIIDERTKEFKEKQ-----ETLKS---TSSEAE 1424
QY 971 YNTTKESL 978

QY 826 INVASNIDROI-----VEGRIEBAKULTSDINVIKNEFKLIESISDLYDKHONGL 878
 Db 1227 ITTSQVDDVDVVIIPIGSEEE-----DYDDLQGVVGEAVTPSVIDNI-----1273
 QY 879 DSHFISFEDISKETNGFRIRFINKETGNSIFITEKEIFSEYATHISKEISNIKDTI--936
 Db 1274 -----LSKIENEYEVLYLKLPLAG--VYRSLKKQLENNVMTF---NVNVRKDIILNS 1317
 QY 937 -----FDNV-NGKLVKVNLDAAHEV-----NTLNSAFFIOSLIEYNTTKESL-SNLS 982
 Db 1318 RFNKRENFKNVLESDDLIPYKDLTSSNVVVKDPYKFLNKRDRKDFLSSVNYIKDSIDTDIN 1377
 QY 983 VAMKVQVYAOLFSTGLNTITDASK 1006
 Db 1378 FANDVLGYKILSEKYKSLDLSIK 1401

RESULT 11
 MYS1_YEAST
 ID_MYS1_YEAST STANDARD; PRT; 1928 AA.
 AC P08964;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Myosin-1 isoform (Type II myosin).
 GN MYO1 OR YHR023W.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=91088308; PubMed=2263482;
 RA Sweeney F.P., Watts F.Z., Pocklington M.J., Orr E.;
 RT "The MyoI Gene from Saccharomyces cerevisiae: its complete nucleotide
 Nucleic Acids Res. 18:7147-7147(1990)."
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=94378003; PubMed=8091229;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
 RA Du Z., Favellio A., Fulton L., Gattung S., Geisel C., Kirsten J.,
 RA Kucaba T., Hillier L.W., Jier M., Johnston L., Langston Y.,
 RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
 RA Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
 RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
 RA Vaudin M.;
 RA "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
 VIII.";
 RT Science 265:2077-2082(1994).
 RL [3]
 RN SEQUENCE OF 1-760 FROM N.A.
 RP STRAIN=S288C;
 RX MEDLINE=88111539; PubMed=3322809;
 RA Watts F.Z., Shields G., Orr E.;
 RT "The yeast MYO1 gene encoding a myosin-like protein required for cell
 division.";
 RL EMO J. 6:3499-3505(1987).
 CC -!- FUNCTION: Required for cell division.
 CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC -!- SIMILARITY: Contains 1 IQ domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X53947; CAA37894.1; --
 DR EMBL; X06187; CAA29550.1; --
 DR EMBL; U10399; AAB68872.1; --
 DR PIR; S46773; S46773.
 DR HSRP; P08799; IAMD.
 DR GeronOnline; 139340; --
 DR SGD; S0001065; MYO1.
 DR GO; GO:0000142; C:contractile ring (sensu Saccharomycetes); IDA.
 DR GO; GO:0007120; P:axial budding; IMP.
 DR GO; GO:0000910; P:cytokinesis; IMP.
 DR GO; GO:0006970; P:response to osmotic stress; IMP.
 DR InterPro; IPR000048; IQ region.
 DR InterPro; IPR001609; myosin head.
 DR Pfam; PF00063; myosin_head_1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SMO0015; IQ; 1.
 DR SMART; SMO0242; MYSC; 1.
 DR PROSITE; PS00096; IQ; 1.
 KW Myosin, Actin-binding; ATP-binding; Coiled coil.
 FT DOMAIN 1 793 MYOSIN HEAD-LIKE.
 FT DOMAIN 794 823 IQ.
 FT DOMAIN 856 1911 COILED COIL (POTENTIAL).
 FT NP_BIND 180 187 ATP (BY SIMILARITY).
 FT DOMAIN 460 529 ACTIN-BINDING (BY SIMILARITY).
 FT CONFLICT 36 36 K -> I (IN REF. 1).
 FT CONFLICT 46 46 I -> T (IN REF. 1 AND 3).
 FT CONFLICT 59 59 V -> S (IN REF. 1 AND 3).
 FT CONFLICT 86 86 L -> F (IN REF. 1).
 FT CONFLICT 330 330 MISSING (IN REF. 1 AND 3).
 FT CONFLICT 343 343 N -> S (IN REF. 1 AND 3).
 FT CONFLICT 421 426 QOAKFI -> TKLSSL (IN REF. 1).
 FT CONFLICT 515 515 D -> S (IN REF. 1).
 FT CONFLICT 529 535 SKGPTG -> ARGHDR (IN REF. 1 AND 3).
 FT CONFLICT 541 541 TD -> V (IN REF. 1 AND 3).
 FT CONFLICT 550 551 TD -> LM (IN REF. 1).
 FT CONFLICT 573 573 R -> A (IN REF. 1).
 FT CONFLICT 582 582 H -> D (IN REF. 1).
 FT CONFLICT 588 599 EYTVGWLKQNK -> NTLWKAGVPT (IN REF. 1).
 FT CONFLICT 599 599 MISSING (IN REF. 3).
 FT CONFLICT 627 632 EKSSA -> GKLLVC (IN REF. 1 AND 3).
 FT CONFLICT 695 695 R -> S (IN REF. 1 AND 3).
 FT CONFLICT 736 742 ENSTTT -> RKNHHD (IN REF. 3).
 FT CONFLICT 756 756 E -> R (IN REF. 1 AND 3).
 FT CONFLICT 773 784 NTLKFFKAGVLA -> ILTVQKLEYWS (IN REF. 1).
 FT CONFLICT 793 794 KL -> NV (IN REF. 1).
 FT CONFLICT 896 896 N -> T (IN REF. 1).
 FT CONFLICT 900 900 N -> NSQITKININITETPOSTYIGERPKVICGN (IN REF. 1).
 FT CONFLICT 906 906 N -> I (IN REF. 1).
 FT CONFLICT 911 911 N -> K (IN REF. 1).
 FT CONFLICT 915 930 NESLNRKTSSETLQ -> RIAIKILKPAINIT (IN REF. 1).
 FT CONFLICT 934 939 DDLVSE -> MTLFL (IN REF. 1).
 FT CONFLICT 951 953 AQN -> RKI (IN REF. 1).
 FT CONFLICT 955 958 EEAH -> KKLD (IN REF. 1).
 FT CONFLICT 1002 1002 S -> C (IN REF. 1).
 FT CONFLICT 1049 1049 L -> D (IN REF. 1).
 FT CONFLICT 1056 1056 C -> S (IN REF. 1).
 FT CONFLICT 1060 1060 M -> I (IN REF. 1).
 FT CONFLICT 1085 1085 A -> E (IN REF. 1).
 FT CONFLICT 1123 1123 V -> C (IN REF. 1).
 FT CONFLICT 1133 1133 L -> S (IN REF. 1).
 FT CONFLICT 1144 1146 KSN -> NLI (IN REF. 1).
 FT CONFLICT 1159 1168 RETKEQKQK -> TRKEEQDKE (IN REF. 1).
 FT CONFLICT 1179 1181 SKI -> ELKV (IN REF. 1).
 FT CONFLICT 1184 1185 LE -> WK (IN REF. 1).
 FT CONFLICT 1188 1204 LSQEISLNOYLNKRISG -> CHRKYLKSLKQKNIR (IN REF. 1).
 FT CONFLICT 1224 1224 P -> S (IN REF. 1).
 FT CONFLICT 1228 1228 E -> Q (IN REF. 1).
 FT CONFLICT 1253 1253 E -> Q (IN REF. 1).
 FT CONFLICT 1311 1323 PKESDINKLMLE -> LTKSLILTNNGAS (IN REF. 1).

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; U67572; AAB99331.1; -
 DR TIGR; MJ1322; -
 DR HAMAP; MF 00449; -; 1.
 DR InterPro; IPR003593; AAA ATPase.
 DR InterPro; IPR003439; ABC transporter.
 DR InterPro; IPR007517; Rad50 zn_hook.
 DR InterPro; IPR003405; SMC C.
 DR InterPro; IPR003395; SMC N.
 DR Pfam; PF04423; Rad50 zn_hook; 1.
 DR Pfam; PF02483; SMC C; 1.
 DR Pfam; PF02463; SMC N; 1.
 DR ProDom; PD000006; ABC transporter; 1.
 DR SMART; SM00382; AAA; 1.
 KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
 FT NP BIND 32 39 ATP (BY SIMILARITY).
 FT DOMAIN 158 849 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1005 AA; 119387 MW; 98BBB48173E788F3 CRC64;

Query Match 4.78; Score 242; DB 1; Length 1005;
 Best Local Similarity 21.28; Pred. No. 0.014;
 Matches 210; Conservative 165; Mismatches 279; Indels 338; Gaps 59;

QY 4 VNKAQLQKVVYKFRQEDSWALLNALEB---YHNMSESSVVEKYLKLDNNLTNY 59
 DB 132 IKQGEIAKFLSLK-PSEKLTAKLIGIDFEKCYQOMGE--IVKYEKLEKIEGELNY 188
 QY 60 LNTYKSGRNKALKKPKFYLTMVELELKNLSPVVEKNLHFIWIGQINDTAINYNQWK 119
 DB 189 KENYEKELNKK-----MQLEKKNKLMWEINDKLN--KIKKEFEDIE-KLENEWE 235
 QY 120 DVNSDYTVKVFYDSNAFLINTLKK-----TIVSATNNTLESFRENLDNPEFD 167
 DB 236 N-----KLLYEK--FINKLEBKRALEKQELKILEYDLNTVVEA-RETLNHKB 285
 QY 168 YNKF-----YKGMELIYKQKHPIDYK--SQIENPEFIDNIKIYLSNEYSKDLE 219
 DB 286 YEKYSLVDIRKIESRLKSHYEDYLTQKLE-----IK-----GDIE 328
 QY 220 ALNKVIEESLKNITANNNDIRNLE-----KFADEDLVRLNQELVERWNLAAASDLIRIS 275
 DB 329 KLEFINKSKYR-----DDINDLTLLNKIKE-----IERVE 361
 QY 276 MLKEDGGVLDVILPGIQDLPKSNKPSITNTSMWIMKLEAIMKYK---EYIPGYTS 332
 DB 362 TIK-----DLL-----ELKNLE-----EIEKIEKYKICECKEYVE 395
 QY 333 KNFMDLDEYVRS-----FESALSSKSKSEIFLDDIKVSPLEVKVAFANNVINOAL 387
 DB 396 KYLEEKEAVEYNKLTLEYITLQEKKSIEK-----NINDLETRINKLLEETKNIDI 447
 QY 388 ISLXDSVCSDLVINQIKRNLNDNLNPSINEGTDENTMTKIFPSDKLASINEDNMFM 447
 DB 448 ESIENS-----LKEIBEKKVVL-ENLQ-----KELGIENSE--IKRL 488
 QY 448 IKITNYLK--VGPAPDVRSTNLSPGVYTGAYQDLMLFQXNSTNHLLEPELRNPEFPK 505
 DB 489 KKILDELKEVEGKPLCKTPI-----DENKMLINQH-----K 522
 QY 506 TKISQTEQETISLWSFNQ--ARAKSQFEYKKGFEFEGALGEDDNDLFAQTVLDKDYVS 563
 DB 523 TQLN---NKYTELEENKIKREIEKIEKLKK-----EIDKEENLKTLYLEK----- 569
 QY 564 KXILSSMKTNRNKEYIHVIVQLQGDKIYSEASCNLFPSKDPYSSILYQKNIGSETAYIVYV 623
 DB 570 -----QSQIEELSLLKKNY- 583
 QY 624 ADASIKEDKYRIYQISNKNENIKLFTIGHCKSEFNTDTFANLDVDSLSSEISTILNLA 693
 DB 584 -KEQLDEINK-----KLSN-----YVINGK-----PVDEILEDIKSQLNFKF 619

QY 684 ADISPKYIEINLLCNMFYSISABETYPKLLIKIDRVSELMPSISQDSITVSANOYE 743
 DB 620 -NFYNOYL-----SAVSYLSNVDDE-----GIENRIKEI-----ENIVSGMKNKE 658
 QY 744 VR--INE--EGKREI---LCHSGKWKINKESSIIKIDISSKEYISFNPKENKIIVKSKYLHE 796
 DB 659 CREELNKLREDEREINLKKLNELKNKEKELI-EIENRRSLKPD-KYKEYLGLUTEKLEE 716
 QY 797 LSTL---LOEIRNNANSSDI---DLEKKWMLTECEI---NVASNIDROI--VEGRIEBAK 845
 DB 717 LANIKDGLLEEIYNTCSKILAIKYNKEDIEIYLNKLEVNKEINDIEBRIYSI- 775
 QY 846 NLTSDSINYIKNEFKLIES-SDSLYDIKHQGLDSDSHFISFEDISKTFENGFRIFIN--- 902
 DB 776 NOKLDEINYNNEEHKKIK-----ELYENKROE-LDNVR-----EOKTEIETG--IBYLKDV 824
 QY 903 -----KETGNSIFITEKEIFSEYATHISK 927
 DB 825 ESKARLKEMSN---LEKEKEKLTKFVEYLDK 853

RESULT 13
 MSPI_PLAF3
 ID MSPI_PLAF3 STANDARD; PRT; 1682 AA.
 AC P19598; Q25921;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
 DE (PMWSA) (P190).
 GN MSP-1.
 OS Plasmodium falciparum (isolate ro-33 / Ghana).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5934;
 RN [1]
 RP SEQUENCE OF 1-1061 FROM N.A.
 RX MEDLINE=8816657; PubMed=3327688;
 RA Certa U., Rotmann D., Matile H., Reber-Liske R.;
 RT "A naturally occurring gene encoding the major surface antigen
 precursor p190 of Plasmodium falciparum lacks tripeptide repeats."
 RL EMBO J. 6:4137-4142(1987).
 RN [2]
 RP SEQUENCE OF 1032-1682 FROM N.A.
 RX MEDLINE=95354793; PubMed=7628566;
 RA Tolle R., Bujard H., Cooper J.A.;
 RT "Plasmodium falciparum: variations within the C-terminal region of
 merozoite surface antigen-1."
 RL Exp. Parasitol. 81:47-54(1995).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 (Potential).
 CC -!- PTM: Merozoite surface antigen contain the sequence of 83 kDa, 42
 kDa and 19 kDa antigens which are the major surface antigens of
 merozoites. The maturation take place during schizont.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; M35727; AAA29715.1; -
 DR EMBL; Y00087; CAA68280.1; -
 DR EMBL; Z35326; CAA84555.1; -
 DR InterPro; IPR006209; EGF_like.
 DR Pfam; PF00008; EGF; 1.
 DR Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1682 MEROZOITE SURFACE PROTEIN 1.
 FT TRANSMEM 1666 1682 MEMBRANE ANCHOR (POTENTIAL).

FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 462 462 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 599 599 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 785 785 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 881 881 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 901 901 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 947 947 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1178 1178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1569 1569 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1682 AA; 192462 MW; C82A1E159948CAD6 CRC64;
Query Match 4.6%; Score 241; DB 1; Length 1682;
Best Local Similarity 19.7%; Pred. No. 0.028;
Matches 206; Conservative 178; Mismatches 376; Indels 286; Gaps 49;
QY 36 NMSSESVVEKYLKLDIINLTNYLYKSGSRNKALEKFEYLTMVELEKNNSLPVE 95
DB 100 NPSDDSSDSDAKYADLKHVQNYLFT-----IKELKYPELFTDTHMLTUCD 147
QY 96 KNLHFIWIGQINDTAINY-INQKDVNSDYTVKVFYDSNAFLINTLKTIVTESATNTL 154
DB 148 -NIH-----GPKYLDIGYEIN-----ELLYKLN-FYFDLLBAKLVNDVANDYC 189
QY 155 E-SFRENLDNPDYNNK-----PYKRMELIYDKQHFIDYKQSEENPEFIINDIKTY 209
DB 190 QIPFNLRKANELDVLKLVFGYRPLDPIKDNVGMEDYIKON-----KTT 236
QY 210 LSEVSKDLKALNKYIEESLNKINTANGNDIENLEK---FADEDLVRLYNOELVERWLA 266
DB 237 IAN-----INELIEGSKTIDQNKADNEEGKKLYQAYDLF-IYNQLOBAHNL 287
QY 267 AASDILRISMLKEDGGVLDVILPGIOPDLFKSINKPDSINTSWEMIKLEAMKYKEY 326
DB 288 SVLE-KRITLKNENIKKLLIEDIKIDAEKPTGVNQILS-----LRLEKSRHEEK 341
QY 327 IP--GYTSK-NFEDMLDEEVQSFESALSSKSEIFLPLDDIKVSPLEVKIAFANNSVI 383
DB 342 IKELIAKTIFENIDRLFTD-PLELEYLYREKKNKVDVTPKSPQTPKSVQIPKVPYENGIVY 400
QY 384 NQALISLKSYSCLVINOIKRY-KILNDNLNPSINEGTDTNTMTKIFSDKLASINED 442
DB 401 PLPLTDIHNLAAD-----NDKNSYGLMNPHTKEKINE-----KIITD-----NKE 442
QY 443 NMFMFKITNLYKVGAPDVRSTINLSGPGVTVGAYQDMLLMFKDNSTHILHLEPLRLNFE 502
DB 443 RKIFI-----NNIKKQIDL-----BEKNIN 462
QY 503 FPKTKISQLTEQITSLWGFNOARAKSQPEEYKKGVEGALGEDDNLDPAQNTVLDKDYV 562
DB 463 HTYEQNKLL-----DYEKSKDYELLEKEFEYEMKFNNNFKD-----VVDKIFS 508
QY 563 SKKLSMKTR-NKEYIH-----YIVOLGDKISYEASCNLFPSKDPYSILYQKNIEGSE 616
DB 509 ARYTANVEKQRYNNKFSNNNSVYNQVKLKALSY-----LEDYS-----LRKGISEKD 557
QY 617 TAYYYV--ADAEIKEDKRIPIQISNKRN--IKLTFIG--HGKSEFTDTTFANLQV- 668
DB 558 FNHYTLTKTGLEADIKKLTAE-----EIKSENKILEKPKGLTHSAN-----ASLEVS 605
QY 669 DLSLSSEIETILNAKADISPKYIEINLLGCNNFYSISAEETYPGK-----LLKI 719
DB 606 DIVLQVQVLLIKKIEDLRK-IEFLKNAQK-KDSIHVPNIYKQNKPEPYLVLVKE 663
QY 720 KDRVSELM-----SISQDSITVSANQYE-----VRINEEGREILDHSGK 760
DB 664 VDLKLEFIPKVDMLKKEQAVLSSITQPLVAASETTEDGSHSTHTLSQSGETEVEETEE 723
QY 761 WINKESIKIDISSKYEYISFNPKENKIIVKSKYLHELSTLLQELIRNANSSDIDLEKKVM 820
DB 724 TVGHTTTVT-----ITLPPKEVKVENS-----IEKSNDSNQSALTKTIVY 763

QY 821 LTECEINVASN-----IDRQIVEGRIEAKNLTS-----DSINYIKNEFKL 861
DB 764 LKLDLBEFLTKSYCHKYLIVSNSSMDQKLELVNTLPENELKSCDRDLLEFNQNNIPA 823
QY 862 IESISDSL-YDLKHQNGLDSDHFIISFEDISKTEGFRIRFINKETGNSIFITEKEI--- 917
DB 824 MYSLYSDMNDLQH-----LFFELYQXEMIVYLHLKKEENHIKKLLEPKQITGT 873
QY 918 -----FSEYATHISKEISNIKTIDFNVNGKLYKK--VNLDAAHEVNTLNSAF 963
DB 874 SSTSSFGNTVTNTAQSATHSNQOOSNASSTNTQNGVSSGPAVVEESHDLTVLSI- 932
QY 964 FIQSLJEYNTTKESLNLSVAMKVOV 989
DB 933 -----SNDLKGIVSLLNLGNKTKV 951
RESULT 14
MSPI_PLAFC
ID MSPI_PLAFC STANDARD; PRT: 1726 AA.
AC P04934;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Merzoite surface protein 1 precursor (Merzoite surface antigens)
DE (EMBL) (P195).
GN MSP-1.
OS Plasmodium falciparum (isolate Camp / Malaysia).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5835;
RN [1]
RP SEQUENCE OF 1-1103 FROM N.A.
RX MEDLINE=86205236; PubMed=3517809;
RA Weber J.L., Leininger W.M., Lyon J.A.;
RT "Variation in the gene encoding a major merozoite surface antigen of the human malaria parasite Plasmodium falciparum."
RL Nucleic Acids Res. 16:1206-1206(1988).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (potential).
CC -!- PFM: Merzoite surface antigen contain the sequence of 83 kDa, 42 kDa and 19 kDa antigens which are the major surface antigens of merozoites. The maturation take place during schizont.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; X03831; CAA27446.1; --
DR PIR; A23386; SAZQGM.
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merzoite; Polyporein; Repeat; Signal; Glycoprotein; Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1726 MERZOITE SURFACE PROTEIN 1.
FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 944 944 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 990 990 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 SQ SEQUENCE 1726 AA; DBAD45FA352BCF3 CRC64;
 Query Match 4.6%; Score 240; DB 1; Length 1726;
 Best Local Similarity 20.4%; Pred. No. 0.032;
 Matches 211; Conservative 172; Mismatches 361; Indels 292; Gaps 52;
 QY 51 DIANNLTNYNTYKSGRNKALKKFKYLTWVLEKLNKSLTPVEKNLHFIWIGGQINDT 110
 DB 154 DLKRVNLYFT-----IKELYPELEDTNMLTCD-NIH----- 189
 QY 111 AINY-INQKVDVSDYTKVYDSNAFLINTLKKTIVESATNTLE-SFRENLDPEFY 168
 DB 190 GPKLIDGYEIN-----ELLYKLN-FYPLLRAKLVNDVCANDYQCPFNKIRANELD 243
 QY 169 NK-----FYKRMELIYDKQHFIDYKSGQIEENPEFIIDNIITYLSNEYSKDLKALNKY 224
 DB 244 LKLVFGYRPLDNKIDNVGMEDYIKKN-----KTTIAN-----INEL 282
 QY 225 IESLNTKITTANNNGNDIRNLEK---FADEDLVRVLYNQELVERWNLAAASDILRISMLKEDG 281
 DB 283 IEGSKTIDQKNADNEEGKKLYQAOYDL-SIYNKOLEAHNLISVLE-KRITDTLKKE 340
 QY 282 GVLVDVILPGIO-PDLFKSINKPDSITNTSWEMIKLEAMKYKEVPGYTSKNFOMLDE 340
 DB 341 NIKELLKINEIKNPPANSNGTNTWLLDKNKKIEHEE--KIKE-IAKTIKENISLFT 397
 QY 341 EVQRSFSAJSSKSKSEIFLPDDIKVSPLEKIAFANNSVINOALISLKSYSCLVI 400
 DB 398 D-PLELEYLREKNKVDVTPKSDQTPKSVQIPKVPNGIVYPLPLTDIHNSLAAD--- 453
 QY 401 NQIKNRY-KILNDMLNSINEGTDFNTMKIFSDKLASINEDNMFMKITYLVKVGPA 459
 DB 454 -NDKNSYGDMLNPDTKEKINE-----KIITD-----NKKRIFI-----NNIK- 490
 QY 460 PDVSTINLSGPGVYTGAYQDLMFKDNSTNHLLEPELNFPPFKTKISQLETSITSL 519
 DB 491 -----KOIDLEKKINHTKEQNKULL 511
 QY 520 WSNFQARAKSOFEBYKGYFEGALGEDDNLDFQNTVLDKDYVSKILSMKTRNKE--- 576
 DB 512 EDYE--KSKDYBELLEKFE-----MKFENN--FDKQVD-KIFSARYVYVVEKQR 558
 QY 577 YIH-----YIVQLOQDKISYEASCNLFSDKDPYSILYQKNIEGSETAYVYV---A 624
 DB 559 YNNKFSSSNNSVYNQVKKKALS-----LEDYS---LRKGISEKDFNHYVTLTKGL 607
 QY 625 DABKEIDKYRIPQISNRYN--IKLTFIG--HGKSEFNTDTPANLDV-DLSSEIETIL 679
 DB 608 EADIKKLT-----EIKSSENKILKRPKGLTHSAN-----ASLEVYDIVKLQVQVL 655
 QY 680 NLAKADISPKYIEINLLGCMFYSISIAEETYPCK-----LLKIKDYRVSLEMP-- 728
 DB 656 LIKKIEDLRK-IEFLKNAQL-KDSIHVPNIYKQNPKEPYLILVKEVDKLKEFIPKV 713
 QY 729 -----SISQSDITSANQYE-----VRINEEGREILDHSGRWINKESITKD 771
 DB 714 KDLMLKEQAVLSSITQPLVAASETTEDGGHSTHTLSQSGTEVTEETEE---TEETV--G 768
 QY 772 ISSKEYISFNPKENKIIVSKYLHELSTLLOEIRNANSDDIDLEKKVLMTECEINVASN 831
 DB 769 HTTITVITLPPKEVKVENS-----IBKNSDNSQALTKTVLKLKDLDEFUTKS 816
 QY 832 -----IDRQIVEGRETEBAKNLTSDSINIK-----NEFKLIESISDSL-Y 870
 DB 817 YICHKIIVNSMSMDQKLELYNLTPEENELKSCDPLDLFNTQNNIPAMYSYDSMNN 876
 QY 871 DLKHQNGLDHSHFISFEDISXTENGFRIRFINKETGNSIFETEKEI----- 917

DB 877 DLQH-----LFFELYQKEMIIYHLKKEENHIKLEBQKQITGTSTSSPGNTT 926
 QY 918 --FSEYATHISKEISNIKIDIFDNGVKLVKK--VNLDAAEHVNTLNSAFFIQSLIEYNT 973
 DB 927 VNTQASATHSNQSQNASSTNTQNGVAVSSGPAVVEESHDPVLVLSI-----SND 978
 QY 974 TKESLSNLSVAMKVQV 989
 DB 979 LKGIVSLNLGNKTKV 994
 RESULT 15
 MSPI PLAPP STANDARD; PRT; 1726 AA.
 ID MSPI PLAPP STANDARD; PRT; 1726 AA.
 AC P50455;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
 DE (PMMSA) (GP195).
 GN MSP-1.
 OS Plasmodium falciparum (isolate Palo Alto / Uganda).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=57270;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89005525; PubMed=3049134;
 RA Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,
 RA Siddiqui W.A.,
 RT "Plasmodium falciparum: gene structure and hydrophobic profile of the
 RT major merozoite surface antigen (gp195) of the Uganda-Palo Alto
 RT isolate."
 RL Exp. Parasitol. 67:1-11(1988).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (Potential).
 CC -!- PTM: Merozoite surface antigen contain the sequence of 83 kDa, 42
 CC kDa and 19 kDa antigens which are the major surface antigens of
 CC merozoites. The maturation take place during schizont.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M37213; AA29611.1; .
 DR InterPro; IPR006209; EGF-like.
 DR Pfam; PF00008; EGF; 1.
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 638 638 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 944 944 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 990 990 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. . .) (POTENTIAL) .
 SQ SEQUENCE 1726 AA; 196174 MW; 5B59CEEP2F9A026 CRC64;
 Query Match 4.6%; Score 240; DB 1; Length 1726;
 Best Local Similarity 20.4%; Pred. No. 0.032;
 Matches 211; Conservative 172; Mismatches 361; Indels 292; Gaps 52;

Qy	51	DINLTDVNILNYTKKSGRNKALKPEKEYLTHEVLELKNNSLTPVPEKMLHPWITGGQINDT	111
Db	154	DLKXRVNTLFT-----IKELKYPFLDTNMLTLCD-NIH-----	189
Qy	111	AINY-INQWKVNSDYTKVFDVSNAFILNTLTKTIVESATNNTLE--SPRENLDNPEFY	168
Db	190	GFKYLIDGVEEIN-----ELLTKLN-FYFDLURAKINDVCANDYCOJPFNLKIRANELOV	243
Qy	169	NK-----FYKRMBIIVDKQKHFIDYKQIBENPEFIIDNIITKLYSNEYSKOLEALNKY	224
Db	244	LKKLVFGYKPLDNIDKVNKMDIYKKN-----KTTIAN-----INEL	282
Qy	225	IEBSLNKITPANGNDIRNLEK---FADBDVLRLYNQBELVERWNLAASDILRLSMLEKQBG	286
Db	283	IEGSKTIDQNKADNEEGKKLYQAOYDL-SIYNKOLEEAHNLISYLE-KRIDTLKQNE	340
Qy	282	GVYLDVJLPGIQ-PDLFKSINKPDSINTNTSWEMIKLEAMKYKEYIPGYSKQNFMDLOE	340
Db	341	NIKELLDKLINEIKNPPANGNTPTNLDDKNKKEEHEE--KIKE-TAKTIKFNIDSLFT	397
Qy	341	EVORSFESALSSKSKSEIFLDDIKVSPLEKIAFANNVSINQALISIKDSYCSDLVI	400
Db	398	D-PLEEYVLRKQKVDYTPKSQOPTKSVQIPKVPYNGIVYPLPTDHNLSAAD----	453
Qy	401	NOIKNRY-KILNDNLNPSINEGTDFNTMTTKIFPSOKLASIGNEDNMMPMIKITNLYKVGFA	459
Db	454	NDKNSYGLMNPDTKEKINE-----KIITD-----NKRKIFI-----NNIK----	490
Qy	460	PVRSITINISGGVYTGAYQDMLLMFKDNSTNHLLEPELRNFPFPKTKISQLEQBITSL	519
Db	491	-----KQIDJEEKINHTEQNKULL	511
Qy	520	WSFNQARAXSQPEEYKGYFEGALGEDDNDLPQNTVLDKDYTSKTLSSMKTRNKE--	576
Db	512	EDYE--KSKDYBELLEKYE-----MKFNNN--FDKQVD-KLPSARYTNVEKOR	558
Qy	577	YIH-----YIVQLQDKISYEASCNLFKDPYSILYQKNIEGSETAYTYVV--A	624
Db	559	YNNKFSSNNSVYVQKLKALS-----LEDYS--LRKGISKDFNHYITLKTGL	607
Qy	625	DABIKEDKVRIPYQISNRN--IKLTFIG--HOKSEFNTDTFANLDV-DLSSEIETIL	679
Db	608	EADIKKLTE-----BIKSENKILEKNFKGLTHSAN-----ASLEWYDVKLVQKVVL	655
Qy	680	NLAKADISPKYEINLLGCMFWSYSABEYTPGK-----LLIKIKDRVSELMP--	728
Db	656	LIRKIEDLRK-IELFLXNAQL-KDSIHVPNIYKQNPKEPYLILVKFVDKLEKIPKV	713
Qy	729	-----SISODSITVSANCYE-----VRINEGKEIILDSKWINKEESIIO	771
Db	714	KDMLKKEQAVLSITQPLVAASETTEDGSHSTHLSQSGETEVTETEER---TEEV--G	768
Qy	772	ISKEYISFPNKENKIIVKSKYLHELSTLLOEIRNANSDDILEKKVMLTECEINVASN	831
Db	769	HTTAVTITLPPKEVKVENS-----IEKSDNDSQALTKTVLWKLDLEPLTKS	816
Qy	832	-----IDRQIVEGRITEEAKNLTSDSINVIK-----NEFKLESISDSL-Y	870
Db	817	YICHKYILVNSMSMDQLEVNLTPLBEBENELKSCDPLDLFNQNNIPAMYSLYDSMNN	876
Qy	871	DLKHQGLDSDHSPISPEDISKTENGFRFRFNKGTGNSIFETEKEI-----	917
Db	877	DLQI-----LFEELYQEMYYLHLKEENHIKLEEQKQITGTSSSPGNIT	926
Qy	918	--PSEYATHISKEISNIKOTIFDNVNGKLVRK--VNLDAAHEVNTLNSAFFIOSLEYNT	973
Db	927	VNTAQATHSNSNQOQSNASSNTTQNGVAVSGPAAVVEESHDPJLTVLSI-----SND	978
Qy	974	TKESLNLNVAMKVQV	989
Db	979	LKGIVSLNLGNKTKV	994

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 16:38:25 ; Search time 14.25 Seconds
(without alignments)
5885.288 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1020

Perfect score: 5189

Sequence: 1 MNLVNAQLQKVMYVFRQ.....ITDASKVELSTALDETID 1020

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Pir1.*
- 2: Pir2.*
- 3: Pir3.*
- 4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5189	100.0	2364	140884	cytoxin L - Clos
2	4101	79.0	2366	2	toxin B - Clostrid
3	4088.5	78.8	2367	2	toxin A - Clostrid
4	2791	53.8	2710	2	alpha-toxin - Clos
5	1435.5	27.7	2178	2	toxin B - Escheric
6	419.5	8.1	3169	2	adherence factor T
7	350	6.7	3225	2	adherence factor T
8	348	6.7	2401	2	phoptry protein -
9	345.5	6.7	1127	2	OSF XSV156 hypothe
10	328	6.3	3724	2	hypothetical prote
11	319.5	6.2	2166	2	hypothetical prote
12	318.5	6.1	3394	2	hypothetical prote
13	316.5	6.1	3255	2	adherence factor T
14	315	6.1	1302	1	surface-located me
15	315	6.1	2269	2	phoptry protein -
16	306	5.9	3335	2	adherence factor T
17	303	5.8	2819	2	conserved hypothet
18	294	5.7	1252	2	reticulocyte-bindi
19	290.5	5.6	4688	2	hypothetical prote
20	287.5	5.5	1447	2	hypothetical prote
21	287	5.5	1979	2	hypothetical prote
22	283	5.5	1306	2	ORF MSV152 probabl
23	283	5.5	1516	2	RAD2 endonuclease
24	282.5	5.4	2829	2	reticulocyte-bindi
25	280	5.4	1939	2	repeat organellar
26	272	5.2	980	2	hypothetical prote
27	272	5.2	1711	2	hypothetical prote
28	267.5	5.2	1121	2	hypothetical prote
29	260	5.0	963	2	conserved hypothet

ALIGNMENTS

RESULT 1

I40884
cytoxin L - Clostridium sordellii
C:Species: Clostridium sordellii
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 15-Oct-1999
C:Accession: I40884
R:Green, G.A.; Schue, V.; Montell, H.
Gene 161, 57-61, 1995
A:Title: Cloning and characterization of the cytoxin L-encoding gene of Clostridium sordellii
A:Reference number: I40884; MUID:95369733; PMID:7642137
A:Accession: I40884
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2364 <RES>
A:Cross-references: EMBL:X82638; NID:G1000694; PIDN:CAA57959.1; PID:G1000695
C:Superfamily: cpl repeat homology
C:Keywords: Cytotoxin

Query Match 100.0%; Score 5189; DB 2; Length 2364;
Best Local Similarity 100.0%; Pred. No. 3.6e-188;
Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MNLVNAQLQKVMYVFRQ	DEYVAILNAL	EEVHNMSSESV	VEKYLKLDINNLT	NYL 60
DB	1	MNLVNAQLQKVMYVFRQ	DEYVAILNAL	EEVHNMSSESV	VEKYLKLDINNLT	NYL 60
QY	61	NTYKSGRNKALKKFK	EYLTMEVLEL	KNSLTPVEKNLH	FTWIGQINDTAIN	YNQWD 120
DB	61	NTYKSGRNKALKKFK	EYLTMEVLEL	KNSLTPVEKNLH	FTWIGQINDTAIN	YNQWD 120
QY	121	VNSDYTVKVFYDSNA	FLINTLKKTI	VESATNNTLES	FRNLNDPEFDY	NKFKRMEIY 180
DB	121	VNSDYTVKVFYDSNA	FLINTLKKTI	VESATNNTLES	FRNLNDPEFDY	NKFKRMEIY 180
QY	181	DQKHFDYKSGQISEN	PEFIDNIIK	YLSNEYSKDL	EALNKYTEESIN	KINTANGNDI 240
DB	181	DQKHFDYKSGQISEN	PEFIDNIIK	YLSNEYSKDL	EALNKYTEESIN	KINTANGNDI 240
QY	241	RNLKFADEDLVRLN	QNLVERWNL	AAASDILRI	SMLKXEDGVLD	VDLPGIQDLPKS 300
DB	241	RNLKFADEDLVRLN	QNLVERWNL	AAASDILRI	SMLKXEDGVLD	VDLPGIQDLPKS 300
QY	301	INKPISITNTSWMT	KLEAIMKYK	EYIPGYTSK	NFDMLDEEVQ	RFSFESALSSKSKSEIF 360
DB	301	INKPISITNTSWMT	KLEAIMKYK	EYIPGYTSK	NFDMLDEEVQ	RFSFESALSSKSKSEIF 360
QY	361	LPDDIKVSPLEVKI	AFANNSVIN	QALISLKD	SYSDLVINQAI	KIKRYKILNDNLNPSINE 420
DB	361	LPDDIKVSPLEVKI	AFANNSVIN	QALISLKD	SYSDLVINQAI	KIKRYKILNDNLNPSINE 420
QY	421	GTDFNTMTKIFD	SKLASISN	ENMMFMKI	TYLTKVGPAD	VRSTINLSGPGVYTGAYOD 480
DB	421	GTDFNTMTKIFD	SKLASISN	ENMMFMKI	TYLTKVGPAD	VRSTINLSGPGVYTGAYOD 480

Db 421 GTDFNTTMMKIFSDKLASINEDNMFMKIITNLYKVGAPDVRGTINLSGPGVYTAYQD 480
Qy 481 LLMFKDNSTNHLPELPELRNFPFKTKISQTEQITSLWSPNARAKSQPEEYKKGYPE 540
Db 481 LLMFKDNSTNHLPELPELRNFPFKTKISQTEQITSLWSPNARAKSQPEEYKKGYPE 540
Qy 541 GALGEDDNLDPQNTVLDKQVSKYKLSMTRNKEYIHYIYVQOGDKISYEASCNLPFSK 600
Db 541 GALGEDDNLDPQNTVLDKQVSKYKLSMTRNKEYIHYIYVQOGDKISYEASCNLPFSK 600
Qy 601 DPYSSILYQKNIEGSETAYYYVADAEIKDYKRIPIYQISNKNIKLTFIGHGKSEFNT 660
Db 601 DPYSSILYQKNIEGSETAYYYVADAEIKDYKRIPIYQISNKNIKLTFIGHGKSEFNT 660
Qy 661 DTFANLVDLSSEITETINLAKADISPKYIEINLLGCMFYSISABETYPGKLLKIK 720
Db 661 DTFANLVDLSSEITETINLAKADISPKYIEINLLGCMFYSISABETYPGKLLKIK 720
Qy 721 DRVSELMPSISODSITVSANQYEVINEEGKREILDHSGKWINKEESIIDKISKEYISF 780
Db 721 DRVSELMPSISODSITVSANQYEVINEEGKREILDHSGKWINKEESIIDKISKEYISF 780
Qy 781 NPKENKIIVKSKYLHELSTLLOEIRNANSDDIDLEKKVWLTECEINVASNIDQIVEGR 840
Db 781 NPKENKIIVKSKYLHELSTLLOEIRNANSDDIDLEKKVWLTECEINVASNIDQIVEGR 840
Qy 841 IBEAKNLSDSINYIKNEFKLIESDSLYDLKHONGLDSDHFI SFEDISKTENGFRIRF 900
Db 841 IBEAKNLSDSINYIKNEFKLIESDSLYDLKHONGLDSDHFI SFEDISKTENGFRIRF 900
Qy 901 INKETGNSIFETEKEIFESEYATHISKEISINIKOTIFDNVNGKLVKKNLDAAEVNTLN 960
Db 901 INKETGNSIFETEKEIFESEYATHISKEISINIKOTIFDNVNGKLVKKNLDAAEVNTLN 960
Qy 961 SAFFIQSLIEYNTTKESLSNLSVAMKVQVYQALFSTGLNTITDASKVVELVSTALDETID 1020
Db 961 SAFFIQSLIEYNTTKESLSNLSVAMKVQVYQALFSTGLNTITDASKVVELVSTALDETID 1020

RESULT 2
S10317
toxin B - Clostridium difficile
C:Species: Clostridium difficile
C:Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text_change 15-Oct-1999
C:Accession: S10317; S21894; S22434
R:Barroso, L.A.; Wang, S.Z.; Phelps, C.J.; Johnson, J.L.; Wilkins, T.D.
Nucleic Acids Res. 18, 4004, 1990
A:Title: Nucleotide sequence of Clostridium difficile toxin B gene.
A:Reference number: S10317; PMID:90326540; PMID:2374729
A:Accession: S10317
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-2366 <BAR>
A:Cross-references: EMBL:X53138; NID:G40442; PIDN:CAA37298.1; PID:G40443
R:Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, M.
submitted to the EMBL Data Library, July 1991
A:Description: Comparative analysis of Clostridium difficile toxins A and B.
A:Reference number: S21894
A:Accession: S21894
A:Molecule type: DNA
A:Residues: 1271-2366 <EIC>
A:Cross-references: EMBL:X60984; NID:G40445; PIDN:CAA43299.1; PID:G40446
R:von Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn, M.
Mol. Gen. Genet. 233, 260-268, 1992
A:Title: Comparative sequence analysis of the Clostridium difficile toxins A and B.
A:Reference number: S22434
A:Accession: S22434
A:Molecule type: DNA
A:Residues: 1791-2366 <VON>
A:Cross-references: EMBL:X60984
C:Genetics:
A:Gene: toxB
C:Superfamily: cpl repeat homology

C;Keywords: cytotoxin
Query Match 79.0%; Score 4101; DB 2; Length 2366;
Best Local Similarity 77.6%; Pred. No. 3.9e-147;
Matches 792; Conservative 106; Mismatches 122; Indels 0; Gaps 0;
Qy 1 MNLVNAQLOKQVYKVFRIQDEYVAILNALBEYHNMSSESVKYLKLOINNTDNYL 60
Db 1 MSLVNRKQLEKQVNRFRTOEDEYVAILDALBEYHNMSSESVKYLKLOINNTDNYL 60
Qy 61 NTYKSGRNKALKKFKELYLMEVLELKNLSLTPVEKNLHFIIWGGQINDTAINYNQWKD 120
Db 61 DTYKSGRNKALKKFKELYLMEVLELKNLSLTPVEKNLHFIIWGGQINDTAINYNQWKD 120
Qy 121 VNSDYTVKVFYDSNAFLINTLKTIVESATNTTLESFRENLDPSFDNKFVKRMEIY 180
Db 121 VNSDYTVKVFYDSNAFLINTLKTIVESATNTTLESFRENLDPSFDNKFVKRMEIY 180
Qy 181 DKQKPIDYKQIENPEFIIDNIIKTVLSNEYSKDLEALNKYIEESINKITANGNDI 240
Db 181 DKQKPIDYKQIENPEFIIDNIIKTVLSNEYSKDLEALNKYIEESINKITANGNDI 240
Qy 241 RNLEKFADEDLVLRYNQELVERWNLAAASDILIRISMLKEDGGVYLDVDILPGIQDLPFS 300
Db 241 RNLEKFADEDLVLRYNQELVERWNLAAASDILIRISMLKEDGGVYLDVDILPGIQDLPFS 300
Qy 301 INKPSITNTSWMTKLEAIMKYKYYIPIYTSKNFMDLDEEVQSFESALSSEKSKSEIF 360
Db 301 IEKPSVTVDFWMTKLEAIMKYKYYIPIYTSKNFMDLDEEVQSFESALSSEKSKSEIF 360
Qy 361 LPDDIKVSPLEVKAFANNVINQALISLSDSYCSDLVINQIKRYKILNDNLNPSINE 420
Db 361 SSLGDMASPLEVKAFANNVINQALISLSDSYCSDLVINQIKRYKILNDNLNPSINE 420
Qy 421 GTDFNTTMMKIFSDKLASINEDNMFMKIITNLYKVGAPDVRGTINLSGPGVYTAYQD 480
Db 421 GTDFNTTMMKIFSDKLASINEDNMFMKIITNLYKVGAPDVRGTINLSGPGVYTAYQD 480
Qy 481 LLMFKDNSTNHLPELPELRNFPFKTKISQTEQITSLWSPNARAKSQPEEYKKGYPE 540
Db 481 LLMFKDNSTNHLPELPELRNFPFKTKISQTEQITSLWSPNARAKSQPEEYKKGYPE 540
Qy 541 GALGEDDNLDPQNTVLDKQVSKYKLSMTRNKEYIHYIYVQOGDKISYEASCNLPFSK 600
Db 541 GALGEDDNLDPQNTVLDKQVSKYKLSMTRNKEYIHYIYVQOGDKISYEASCNLPFSK 600
Qy 601 DPYSSILYQKNIEGSETAYYYVADAEIKDYKRIPIYQISNKNIKLTFIGHGKSEFNT 660
Db 601 DPYSSILYQKNIEGSETAYYYVADAEIKDYKRIPIYQISNKNIKLTFIGHGKSEFNT 660
Qy 661 DTFANLVDLSSEITETINLAKADISPKYIEINLLGCMFYSISABETYPGKLLKIK 720
Db 661 DTFANLVDLSSEITETINLAKADISPKYIEINLLGCMFYSISABETYPGKLLKIK 720
Qy 721 DRVSELMPSISODSITVSANQYEVINEEGKREILDHSGKWINKEESIIDKISKEYISF 780
Db 721 DRVSELMPSISODSITVSANQYEVINEEGKREILDHSGKWINKEESIIDKISKEYISF 780
Qy 781 NPKENKIIVKSKYLHELSTLLOEIRNANSDDIDLEKKVWLTECEINVASNIDQIVEGR 840
Db 781 NPKENKIIVKSKYLHELSTLLOEIRNANSDDIDLEKKVWLTECEINVASNIDQIVEGR 840
Qy 841 IBEAKNLSDSINYIKNEFKLIESDSLYDLKHONGLDSDHFI SFEDISKTENGFRIRF 900
Db 841 IBEAKNLSDSINYIKNEFKLIESDSLYDLKHONGLDSDHFI SFEDISKTENGFRIRF 900
Qy 901 INKETGNSIFETEKEIFESEYATHISKEISINIKOTIFDNVNGKLVKKNLDAAEVNTLN 960
Db 901 INKETGNSIFETEKEIFESEYATHISKEISINIKOTIFDNVNGKLVKKNLDAAEVNTLN 960
Qy 961 SAFFIQSLIEYNTTKESLSNLSVAMKVQVYQALFSTGLNTITDASKVVELVSTALDETID 1020
Db 961 SAFFIQSLIEYNTTKESLSNLSVAMKVQVYQALFSTGLNTITDASKVVELVSTALDETID 1020

Qy	600	KDPSYILYQXNTEGSETAYYYVADAEIKEDKYRIPQYISNKRNIKUTFIGHGKSEFN	659
Db	601	KNPYDSILFORNIETSEVAYYNNPTDSEIQEIDKYRIPDRISDRPKIKLUTFIGHKAEPN	660
Qy	660	TDTFANLVDVDSLSSETITNLAKADISPKYIYNLLGCMNPSYSISAEETYQCKLLKI	719
Db	661	TDTFAGLVDVDSLSSETETAIGLAKEDISPKSIYNLLGCMNPSYSYNVVEETYQCKLLRV	720
Qy	720	KDQVSELMPSISQDSITVSAQYVEVEINPEGKREILDHSGKWINKBESIKDISKEYIS	779
Db	721	KDQVSELMPSMSQDSITVSAQYVEVINSEGRRELLDHSGEWINKBESIKDISKEYIS	780
Qy	780	FNPKENKIIVKSXYLHSLTLLQEIENNANSSDIDLEKKVMLTECINVASNIDROIVEG	839
Db	781	FNPKENKIIVKSXNLPESLTLQEIENNNSSDIELEKKVMLAECINVISNIEQVVEE	840
Qy	840	RIEAKNLTSDSINYIKNEPKFLTESDSLYDLKHQGLDSDSHFISFEDISKTENGFRIR	899
Db	841	RIEAKSLTSDSINYIKNEPKLIESIEALCDLKQONELEDSDHFISFEDSETDEGFSIR	900
Qy	900	FINKETGNSIFITEKEIISEYATHISKEISNIKDIFDNVAGKLVKKNLDAAEVNTL	959
Db	901	FINKETGESIFVETEXTIIPSEYANHITEEISKIKGIFDTVNGKLVKKNLDTTHEVNTL	960
Qy	960	NSAFFQSLTEYNTTKESISNLNVAMKVQVYAOLFSTGLNTIDTDAKXVVELVSTALDETI	1019
Db	961	NAAFFQSLLEYNSKESISNLNVAMKVQVYAOLFSTGLNTITDAAKXVVELVSTALDETI	1020
Qy	1020	D 1020	
Db	1021	D 1021	

RESULT 4

A37052
toxin A - Clostridium difficile
C:Species: Clostridium difficile
C:Date: 31-Jan-1992 #sequence revision 31-Jan-1992 #text_change 24-Sep-1999
C:Accession: A37052; A60991; S21897; S22437; S08638
R:Dove, C.H.; Wang, S.Z.; Price, S.B.; Phelps, C.D.; Lyster, D.M.; Wilkins, T.D.
Infect. Immun. 58, 480-488, 1990
A:Title: Molecular characterization of the Clostridium difficile toxin A gene.
A:Reference number: A37052; MUID:90129305; PMID:2105276
A:Accession: A37052
A:Molecule type: DNA
A:Residues: 1-2710 <DOV>
A:Cross-references: GB:M30307; NID:G144925; PIDN:AAA3283.1; PID:G144926
R:Wren, B.W.; Clayton, C.L.; Tabagchali, S.
FEMS Microbiol. Lett. 70, 1-6, 1990
A:Title: Nucleotide sequence of Clostridium difficile toxin A gene fragment and
A:Reference number: A60991
A:Accession: A60991
A:Molecule type: DNA
A:Residues: '1', 1894-1899, 'K', 1901-1910, 'Y', 1912-1919; 2054-2074; 2096-2116; 2138-
A:Cross-references: GB:X17194
R:Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sau-
submitted to the EMBL Data Library, July 1991
A:Description: Comparative analysis of Clostridium difficile toxins A and B.
A:Reference number: S21894
A:Accession: S21897
A:Molecule type: DNA
A:Residues: 1-154 <EIC>
A:Cross-references: EMBL:X60984
R: von Eichel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.;
Mol. Gen. Genet. 233, 260-268, 1992
A:Title: Comparative sequence analysis of the Clostridium difficile toxins A and
A:Reference number: S22434; MUID:92293124; PMID:1603068
A:Accession: S22437
A:Molecule type: DNA
A:Residues: 1-92 <VON>
A:Cross-references: EMBL:X60984
A>Note: The four fragments shown in reference A60991 correspond to four types o
g with repeats ordered ABCCCDABCDABACCCDABCCDABCCDABCDABC

RESULT 3

S70172

toxin B - Clostridium difficile
C/Species: Clostridium difficile
C/Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 15-Oct-1999
C/Accession: S70172; S44271
R/von Eichel-Streiber, C.; Meyer zu Heringdorf, D.; Habermann, E.; Sattngen, S.
Mol. Microbiol. 17, 313-321, 1995
A/Title: Cloning in on the toxic domain through analysis of a variant Clostridium difficile
A/Reference number: S70172; MUID:96079281; PMID:7494480
A/Accession: S70172
A/Status: preliminary; nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 1-2367 <VON>
A/Cross-references: EMBL:Z23277; NID:G761713; PIDN:CAA80815.1; PID:G761714
A/Experimental source: isolate 1470
R:Sattngen, S.; von Eichel-Streiber, C.
submitted to the EMBL data Library, July 1993
A/Description: Cloning and sequencing of an Clostridium difficile toxin B mutant.
A/Reference number: S44271
A/Accession: S44271
A/Molecule type: DNA
A/Residues: 1-1323, 'N', 1325-2367 <SAR>
A/Cross-references: EMBL:Z23277
C/Superfamily: cpl repeat homology
C/Keywords: cytotoxin

Query Match 78.8%; Score 4088.5; DB 2; Length 2367;
Best Local Similarity 77.1%; Pred. No. 1.2e-146;
Matches 787; Conservative 114; Mismatches 119; Indels 1; Gaps 1;

QY 1 MNLVKALQKQVYKFRQEDYVAILNALEYHNMSSVVEKYLKLDINLVNLY 60
DB 1 MSVLNQKLEKMANVFRVQEDYVAILDALEYHNMSENVTVEKYLKLDINSLDTYI 60
QY 61 NTYKSGRNKALKKFEYLTMVELEKNNLSLTPVENLHFIWIGGOINDAINYNQWKD 120
DB 61 DTYKSGRNKALKKFEYLTMVELEKNNLSLTPVENLHFIWIGGOINDAINYNQWKD 120
QY 121 VNSDYTVKYFYDSNAFLNLTAKTIVESATNTLTFSPRENLPDEFYKFKRMEIYY 180
DB 121 VNSDYNVNFYDSNAFLNLTAKTIVESATNTLTFSPRENLPDEFNHTAFKRWQIYY 180
QY 181 DKQKHFDYKQSIENPEPIINIKTYLSNYSKDLEALNKYIEBSLNKITTANNNDI 240
DB 181 DKQQNFINTYKAKENPDLIIDIVKTYLSNYSKIDELNAYIEBSLNKVTENSNDV 240
QY 241 RNLEKPAEDLVRLNQELVERNLAAADILISMLEDGGVYLDVILPGIQPDLKFS 300
DB 241 RNFEFKTGVFVNLVQESVERVNLKAGADILKAVILKNTGGVYLDVDMFGHLPDLFKD 300
QY 301 INKPDSTI-TNTSWMIKLEAMKYKEYIPGYTSKNFDMLEDEEVQSFESALSSKSKSEI 359
DB 301 INKPDVKTAVDWEQWLEAMKKEYIPEYTSKHTDLDVEVQSFESVLAASKSKSEI 360
QY 360 FLPLDDIKVSPLEVKAFANNSVINQALISLKDYSDDLWINQIKRYKILNDLNPSIN 419
DB 361 FLPLGDIENSPLEVKAFAGKSIINQALISAKDSYCSDDLKIQIRYKILNDLTGPIIS 420
QY 420 EGTDENTWKIPDCKLASISNEQNMFMKITYNLKVGPAQVRSNTINLSGGVYTGAYQ 479
DB 421 QGNDFTNTMNFSGSGLAANEENISFIKIGSLRVGFPEANTTITLSGPTIYAGAYK 480
QY 480 DLLMFKDNSTNTHLPELNPFPFKTKISQLTEQETISLWSNQARAKSQFEYKGYF 539
DB 481 DLLTFFKMSIDTSLSELNPFEPFKVNIQSATEQKNSLWQFNERRAKIQFEYKGYF 540
QY 540 EGALGEDDNLDPQNTVLDKDYKSKITLSNKTNRKEYIHVIYVQLOGDKTISYASNLPS 599
DB 541 EGALGEDDNLDPQNTVLDKDYKSKITLSNKTNRKEYIHVIYVQLOGDKTISYASNLPS 600

Qy 872 LKHONGLDGSHFISPEDISKTENGFRIRFNKGTGNSFIETETKEBIPSEVATHISKESIN 931
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 876 FYCKNINSNMIIILFDIIEKDYNNVKLANKITGETSVIKTYSDSLWNFNTKYKKIVDD 937
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 932 IKOTIPDNVNGKLKVYNLDAAEHVNTLSNAFTIOSIEYNTTKESLSNI SVMAKVOYYA 991
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 938 IKGIIVDINGEFIKKADFBIBONPFSLLNSAMLMLQLDIDYKPYTEILTNNMNTSLUKOYA 997
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 992 QLFSTGLNTITDASKVVVELSTALD 1016
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 998 QIFOLSIGAQEATEIVTIISDALN 1022
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
toxin B - Escherichia coli plasmid p0157
C/Species: Escherichia coli
C/Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 17-Nov-2000
C/Accession: T00296; T42195
R/Making: K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C.; Kubota,
S.; Shinagawa, H.
DNA Res. 5, 1-9, 1998
A/Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic
E.coli O157:H7 strain EDL933; PMID:9628576
A/Reference number: Z14127; MUID:98290540; PMID:9628576
A/Accession: T00296
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-3169 <WAK>
A/Cross-references: EMBL:AB011549; NID:G4589740; PIDN:BAA31815.1; PID:G3337056
A/Experimental source: strain EHEC O157:H7, substrain RIMD 0509952
R/Burland, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Sofia, H.J.; Blattner, F.R.
Nucleic Acids Res. 26, 4196-4204, 1998
A/Title: The complete DNA sequence and analysis of the large virulence plasmid of Esche-
ria coli O157:H7 strain EDL933; PMID:9722640
A/Reference number: Z22068; MUID:98391744; PMID:9722640
A/Accession: T42195
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-236,'D',238-1887,'I',1889-3169 <BUR>
A/Cross-references: EMBL:AF074613; PIDN:AAC70163.1
A/Experimental source: strain EDL933; serotype O157:H7
C/Genetics:
A/Gene: toxB
A/Genome: plasmid p0157
A/Note: L7895
C/Superfamily: Escherichia coli probable cytotoxin
C/Keywords: cytotoxin

Query Match 8.1%; Score 419.5; DB 2; Length 3169;
Best Local Similarity 20.8%; Pred. No. 4.1e-08;
Matches 256; Conservative 202; Mismatches 446; Indels 325; Gaps 55;

Qy 14 YVKRI---QEDEVAILNAL--FEYHNMSESSVVEKYLKDKDINNLTNDYNTYKISGR 68
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 160 YIKRTKGAEDQTTTTQSIIINELLNGVDNRITI-PFKISELNLIHSVENMOIKNSR 218
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 69 -----NKALKKFKE---YLTMEVLELKNNSLTPVEKNLHFPIWIGG 105
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 219 KGIBLVKQGELLSLNLNVNKGNSQLSDNASKIINLLGIEYQSHKV-DIEFFIHAVVAG 277
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 106 QINTAINYNQWKVNSDYTVKVFYDSNAF-----DLINKKT----- 144
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 278 APPDNTFSYIAFLNTVKDYTYLLWDPNFAAKFGSKILKNIAMYAIIMRLRNPHLA 337
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 145 -----IVESATNWILE--SPRENLDPEFDY-----NKEYKRMELIIDYKQHF 186
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 338 EEMNEVLKIQIQNETIEFKETRERIKLENRYKSUTSETKEFPNVFLESIMGMQNY 397
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 187 IDYKKSQIEENPEFI----IDNIHK--TYSNVEYSKDEALNKYIEBSLNKITANNNG- 238
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 398 FTYCISNCISNTDDISRDLDFTNVLKSPFQNDFPKSTEKKNKRDIIDLKWTISOQFGDR 457
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 239 ----DIRNLEKFADEDLVRLYNQBELVERWNLAASDILRI SMLKEDGGVYLDVDPGIC 294

Db 458 PQLRDINTLSFFKKPQDYFFYQOEMLRWYAAASDQVRINILKEYGGIYTDITLIPAYS 517
Qy 295 PDLFKSINKPDSINTNTSWEMIKLBAIMKYK--EYIPG--YTSKNFMDLDEVQSRFESAL 350
Db 518 DKVQQLNE-KSDDKRRFFEDLKLRIISSEILSLIKGEKYSIKH-DGLDDETTINQLNNIL 575
Qy 351 SSKSDKGEIFLDDIKVSPLEVKIAFANNSVI-----NOALISLKDSY 394
Db 576 -SEIEK---LTIDY-PPKPVETKVRDTPKIPKRYQKWTENTWIRGNNNFMTLHGSK 629
Qy 395 CSDLVINQIKNRYKILNDNLNPSINEGTDFNTTKIFSDKLASINB-----DN 443
Db 630 CIDFILSQKQKYLEL-QRIDNTSYNNLFYTT-----EDLKSINNVAIGIPAKKYLEH 683
Qy 444 MFMPIKNTYIKVGFADPVRSTINLSP-----
Db 684 GLP-----SEYRQDGTPIYVWSTNLISGPDIMRQMKYKSLGRIGEVHTKDNKLSVNF 739
Qy 472 -GVYTAGYQDLMF---KDNSTNTHLLEPE-----LRNFEPPKTKISQLT---EQEIT 517
Db 740 LGVTASSKNKNSFNWLPVSGVINDTPDESSWAVRNNDINKILPEKINCHVPEKLPT 799
Qy 518 SLWFSNQARAKSQPEYKKGFEFEGALGEDDNLDPQNTVLDKDYVSKKILSSMKTRKEY 577
Db 800 SLY-----YEIDRSFFQGW-----DNKSIKHVTEINKDLI--KDNILLTSSNID 843
Qy 578 IHYIVOLGDKISTEASCNLSPKDPYSILYQKNIEGETAYYYVADAKIKIDKTRIP 637
Db 844 VKLLIKL--DRELVASISKI-----DNPLALRSIRTLQLQLANVYTSNTPEPENTINFI 895
Qy 638 YQISNKN-----IKLTFIGHGKSEFNTDFANLVDSS-----LSSEIE 676
Db 896 YDFVRKQDQLLSAFL-----FSRNDATKLIWVNSMKNVFLREVIS 941
Qy 677 TILNLAADISPKEYTEINLLGOMFYSISAEETVPGKLLIKIDRVSELMPSISQDSI- 735
Db 942 CVLRSKKVD--SYINEN-----KNLSKEDA--GALRDYAKLKKELFSMLDDGDK 989
Qy 736 -TVSANQYEVRAINEGKEILDHGKWKINKEESIKDISKEYISFNP-----782
Db 990 KIITTNAY--IKERDKL-----SGIIVNIENSIIGHESPDITRSNHEWGDLSLTVQF 1041
Qy 793 KENKIIVSKYILHELSTLQIRNANSDDIDLEKKYMLTECEINVAS-----NIDRQIV 837
Db 1042 KKFEPYKVE--LSSAKSFDDIKNK-YITDPETKRNVLHQLDSIKERIAFLDISHAY 1099
Qy 838 EGRIEEAKNLTSDSINYIKNEFKLIESI-----SDSLYDLKHONG 877
Db 1100 PGLSLEKLQLS-----GYVFSINIIEVLLASYGVSGHYGVVYPAPSKLLELRHT 1155
Qy 878 LDDSHFISFEDISTENGFRIRINKETGNSI---FIETEKEIFSEYATHISKEISNID 934
Db 1156 KSNSEMI--EKITP---YVVDILSDNSVNLRPPLSEQKKILNDIKLEISKVS---E 1206
Qy 935 TIFDNNVGNKLVKNLDAHEVNTLNSAFFTCQLIEYNTTK-----ESLSNLSV---983
Db 1207 QYFMKLTEQSSVIGIKYSVDFRNYENLFLSLPINQNLTLFFMYRYPEMLYDIHIGIE 1266
Qy 984 --AMKVQVYQAQFSTGLNTITDASKVVEL 1010
Db 1267 NKANREFIYKFSNLNLDLINDERVLNL 1295

RESULT 7
DB1702
adherence factor TC0439 [imported] - Chlamydia muridarum (strain Nigg)
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: DB1702
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000

A>Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: AB1500; MUID:20150255; PMID:10684935
A:Accession: DB1702
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3225 <1E7>
A:Cross-references: GB:AB0002312; GB:AB000160; NID:G7190482; PIDN:AAF39293.1; PID:G719048
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0439

Query Match 6.7%; Score 350; DB 2; Length 3225;

Best Local Similarity 19.7%; Pred. No. 1.8e-05;

Matches 255; Conservative 207; Mismatches 406; Indels 428; Gaps 64;

Qy 2 NLNVAQIQKQVYKFRIOEDYVAIL-NALIEYHNMSE-----SSVVEKYLKLDINN 54
Db 196 NILEKLSITQ-----EQIKHNLNLSNKEALNRSDLNKRNRKGSLLVQAAILD--E 249
Qy 55 LTDNYLNTYKKSGRNALKKFKXYLTMEVLEKNNSLTPPEKNLHFFIWIQOINDTAINY 114
Db 250 ILSQTKSTEERASNSVMTTIKKEFTSHRV-----PVEKNIHGIWAGSPPEGTDEY 300
Qy 115 INQWQVNSDYTVKVYDSNAF-----LNTLKKTIVESATNTLTLESFRENLDNDFYNK 170
Db 301 IKLFLHTYPEFSFLPWVDKTAAGAAKFSSTLKRIAFDAVN-----SLREATPEPVKQFVQ 356
Qy 171 FRKMEIILYKQKFEFIDYKSOIENFEFIDNIIKTYLSNEYSKDL-----218
Db 357 RYDK-LKKSYSRTSDFDE--KORUSEQVELVDNY-----NKFSKEIQSNPDLVLLHEM 407
Qy 219 -----EALNKY-----EESLNKI 232
Db 408 ITIQSPFNYCOLKGVGALTDETREIYLEKLVKVEEDLSHYKETIKENKESIEKLVEI 467
Qy 233 TANG-----NDIRNLEKFADEDIARLYN--QELVERWNLAAASDIILRIMLKEDGGY 284
Db 468 NDSTGRERWIKDIRLAKSL--QDLTNSYNYETEMLLSNVYAAATDQLRMTYMLKEYGGIY 525
Qy 285 LQVDILPGIQLPKSIN-----KPSITNT-SWEMIKL-----EAIMKYEIIPGT 331
Db 526 TDLDMPOYSQVQLKIMDVGGSRFFEDHKLRTLSFAALKLGSGKQTTVFEAKKAMT 585
Qy 332 SKNFQMLDE-EVQRFESALSSKSKSIFLPLDDIKV---SPLEVKI---AFANNSVI 383
Db 586 LPTFTLQDQKSEIIFKLETOAKSLIFQPMQVTVVRDPMPIQLQRHKWGTGNVRL 645
Qy 384 NOALISLKDSYCSDLVINQIKNRYKILNDNLNPSINEGTDFNTTKIFSDKLASINEDN 443
Db 646 NGLMMAHKDSAVDAVIARQAAAYDEM--ALRQNVVSGEFFRSL-----GDLEHVNREKN 699
Qy 444 M-MFMIKITNYL-----KVGFPADPVRSTINLSGPGVYTGAYQDLM--FKDNSTNI 491
Db 700 IGYLAK--NYLGGSLFPDFRQDSVIPGAISTLGISGP-----DIIMDTMSDYFTNL 749
Qy 492 -----HLEPELRN-----PEFPKTKISQLT-----EQEIT---SLWSFNOA 525
Db 750 GPVGEDFLYEGKLGAAFLGAYCAQKPKGELTYDNLHPLSLGANDVTPADASTCETRQ 809
Qy 526 RAKSQF-----EYKKGYPEGALGEDD-----NLDFAQ-----NTVLDK 559
Db 810 HCAEELLSDSISSEHPKGIERRVNPNDPFSKLSKKAQGLSSDFADLLPRFNLLIES 869
Qy 560 DYVSKKILSSMKTRNKEYIHYIVQLGDKISYEASCNLSPKDPYSIILYQKNIEGETAY 619
Db 870 SALDIHTLSAL--DRDIQHLFTKVQ-----KDPVASVA-----900
Qy 620 YYYVADAETKEIDKVR-IPYQISN-----KKNIKLTFIGHGKSEF---558
Db 901 ---VFSLQQLAEMIRAIPEFFIRNVQVHILPEAQAHFEADWKAQIQLYLSHPQTEVINY 957
Qy 659 ---NTDTFANLVDSSLSSIEIILNLAADISP-----KYIEINLGC-----NMF 701

Db 958 SSTHTQIVFGKOLLAVARAAKSL-MSDHPSLTSLTSYLKYKTOHSHGLVLTFFQEDOFF 1016
QY 702 SYSIS-ABEYTPGKLLKIKDRVSELMPSISQDSITVSANQVEVRAINEEGKREILDHS-G 759
Db 1017 ELMDVIAEPELPKQLKIEQVN-----SGLYSHVEHSLG 1052
QY 760 KW--INKEE-----SIKDISKEYISFNPXENKIIVKSKY-----LHEL 797
Db 1053 EMLKLSKERKSKFLKILAKETFEEREDSQOQHTWFEELYEKHQRVQDPKAKIOEL 1112
QY 798 STLLQIRNANSSDID-----LEKKVMTCEINVASNDRQIVGRIEAEAKNLTSD 850
Db 1113 ITVFOE-SORVOAQDIDTYFAHKPFYQDLMDGYAFEDISVITKYL-----LASD 1161
QY 851 SINVIKNEFKLIESDSLYD-LKHQGLDDSHFISFEDISKTENGFRIRF--INKETGN 907
Db 1162 GVGSIITDPIPPPKQIDAMKQSLG-----EDFGSLHYTLQVYDWLSKET-N 1211
QY 908 SIFPIETEK-----EIFSEYATH-----ISKEISNIKDTIFDNV-- 940
Db 1212 SVTSEQAKQKLPKQLEKLEGVYTHDLLIPPIDGVSALGLFSTEEGKVSQVRLTSTAP 1271
QY 941 -----NGKLVKKNVLDAAHEVNTLNSAFFIQSLIBYNTTK 975
Db 1272 GVPNSASYAMTSYLYGLFLITKDIQSGRLTHEI---LKERLQYGGAYFI-----N 1319
QY 976 ELSLSNLSVAM---KVQVY---AQLFSTGLNTITDAS 1005
Db 1320 ESKIDVLLALSRKKAQSLIDAKHALTGFSPFSEAS 1355
RESULT 8
T28676
rhoptry protein - Plasmodium yoelii (fragment)
C;Species: Plasmodium yoelii
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C;Accession: T28676; A45521
R;Sinha, X.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A;Title: Comparison of two members of a multigene family coding for high-molecular mass
A;Reference number: Z20507; MUID:97077455; PMID:8920022
A;Accession: T28676
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2401 <SIN>
A;Cross-references: EMBL:U36927; NID:g1041784; PID:g1041785; PIDN:AAB41263.1
R;Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A;Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple co
A;Reference number: A45521; MUID:91101660; PMID:2270106
A;Accession: A45521
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 2260-2401 <KEE>
A;Cross-references: GB:M34281
Query Match 6.7%; Score 348; DB 2; Length 2401;
Best Local Similarity 20.9%; Pred. No. 1.4e-05;
Matches 262; Conservative 207; Mismatches 406; Indels 377; Gaps 66;
QY 6 KQALQKQVYKPRIQSDVEVAINALAEVYN--MSESVVEKYLKLNINLTQNYLNTY 63
Db 118 KDTMKKIIILLIQYM-NEFGKLDAMTKLKNEGISQKVFVNNQIKQKFKSTYDE----- 171
QY 64 KSGRNKALKKFK--EYLTWEVL-ELKNNSLFPVEKNLHFVIGQINDTAINYNQKD 120
Db 172 KKEGFESSLELAKNWEKKLEIITELKKNKEETVQDLKIRELIKQKD-----IIEEQKI 227
QY 121 VN-----SDYVYKVFYDSNAPLNTLTKTIVESATNTLSTFRENLDNPFQYNK 170
Db 228 VNDLKLELNKKNKEITEKTEYTKAV---DLKKEIKD--NVYID--ELAKEPPYQITK 279
QY 171 FYRKRMEIHYDKQHFIDYKQIEE--NPEFII---DNI-----IKTYLSNEY 214

Db 280 YIEKNEIYNTIKSDFDKIYVGDIQLYNMFVQESNIEHTEILTKTKIDNVY 339
QY 215 S-----KOLEALNKYIEESLNKITANG-----NDIRNLEK----- 245
Db 340 NNIQNMETETVVKSHLKNITNNKLSSETLIDIIKYIYEITNELNKLTEDPKNEKGLSNK 399
QY 246 ---FADBDL-VRLYNOLBVE---RWNLAAASDILRISMLKE-DGGVYLD--VILPQIQP 295
Db 400 IDEVAKENVOLNVYKSNILIEIKHYN-----DQINIDNIKEAKAQNYDQFKHMTIPP 454
QY 296 DLFKSNKPSISNTNTSWEMIKLEAIMKYEYIPGY--SKNPFMDLDEEVQSFESALSKSD 355
Db 455 NEMK-YQKP---SIEIKMKDEFLSKVNY-----NDFDKVYKKEVEHKNFTELTN 503
QY 356 KSEIFLPDDIKVSPLEVKIAPANNVINOALLSLKDSY-----C--SDL 398
Db 504 K--IKTEVSDEEIKKYNKEN-DSKSLINETKKSIEEYQNIINTLKKVDDYIIVCLNTNE 560
QY 399 VINQIKRYKILNDNLNPSINEGTDFTNTWKIFSDKLASI----- 438
Db 561 LITNCHNRQITLQKLNQNIKTIKETNSIDKIITDKFENILTDKKTETKFTGLSLNNH 620
QY 439 -----SNEDNMKFM-----IKITNYLKVGFAPDVRSTINL 468
Db 621 ESNNKELLTYFYDLKANLGNKNENMLYKQFNEKEKAVEDIKKNV-----DINKIVSN 673
QY 469 SGPQVYTGAYQDILMFKDNSTNIHLLEPELRNEFPFKTKISQUTEOBITSLWGFNOARAK 528
Db 674 IEITIVTSIYN--INBDTENEIG-----KSTELLNTKYLEKVKANVTNL----- 715
QY 529 SQPEEYKKGYPGALGEDDNLDFQAQTVL--DKDYVSKILSSMKT-----RNKEIYHYI 581
Db 716 NEIKBKLDYDFQDFGKERNIKYPDENKIKNDITLNQKIDKISLETITELIKKXSE--NHI 773
QY 582 VOLQG--DKISYEASCNLPSKDPVSSLYQKNTEGSETAYYYVADAIEKIDK--YRIP 637
Db 774 DEIKGQIDKLUKVPKNTWNEPD-----KEIEKKIENIV 807
QY 638 YQISNKNRIKLTPIGHGKSEFNTDTFANLDVDSLSEIETILNLAKADISPKYIEINL-- 695
Db 808 EKIDKKNI-----YKEIDKLLNEISKIEN-DKTSLE-KLKNNILSY 847
QY 696 ---LQGNMF-----SYSISAEETYPGKL-LLKIKDRVSELMPSISQD----- 733
Db 848 GKSIG-NLFLOQIDEEKKAEHTIKAMEAYIDDLNIIKKKSQBIKEMNINMDIKMDIHK 906
QY 734 ---SITVSANQYEV-----RINEEGKREILDHSGKIN--KEESIIDKDISK----- 775
Db 907 EMKALNISHDDYKIYHTTSKNHEKISDIRKNSLKIIQDFSEESIINDIKKELEKNVLES 966
QY 776 -----EYIS-----FN-PKENKIIVKSKYLHELSTLLQEIERN-----ANSSDI 813
Db 967 QNNTDINQVLSKIENIYNILKLNKI---KKIIDKVEYDEIEKNNKINAELSSEKI 1023
QY 814 --DLEKKVMTCEINVASNIDRQIVGRIEAEAKNLTSDSIN-----YIK----- 856
Db 1024 ITOLKSSSLKEQSKIKSIIDDNVYSECIKNITNLKTYIVNEKNINITYFKNAEYQN 1083
QY 857 -----NEFKLIESISLSYDLKHQGLD--DSHFISFEDISKTENGFRIRFINKETG-NS 908
Db 1084 VSLNFNNEIADTKSQYLNIKQNGTNTDYNIKELKEHKKSKSNVYK-----DEAGKNT 1138
QY 909 IFIETEKEIFESEYATHI-----SKEISNIKDTIFDNVNGKLVKKNVLDAAHEVNT-- 958
Db 1139 QEIKONKELFEKEQEVTVLNNKYAVELKNKFDKT-KNVSEQIIEIK--DAHNTFTSQ 1195
QY 959 -----LNSAFFIQSLIEIYNTKESLSNLSVAMKVQVYAQLFSTGLNTITD 1003
Db 1196 ADKSEKQWBIKNEQRIEDEVAKNKSNAI-LDIQLSVEPFIKFLKID 1246
RESULT 9

T28317
ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C:Species: Melanoplus sanguinipes entomopoxvirus
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C/Accession: T28317
F:Afonso, C.L.; Tulman, B.R.; Lu, Z.; Oma, B.; Kurish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A:Reference number: Z20484; MUID:99102612; PMID:9847359
A:Accession: T28317
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1127 <AFO>
A:Cross-references: EMBL:AF063866; NID:g4049647; PIDN:AAC97677.1; PID:g4049717
C:Genetics:
A:Note: MSV156

Query Match 6.7%; Score 345.5; DB 2; Length 1127;
Best Local Similarity 21.9%; Pred. No. 6.8e-06;
Matches 248; Conservative 167; Mismatches 368; Indels 349; Gaps 54;

QY 9 LQKVVYKFIQDEYVAIINALBEYHNMSESSVVEKYLKLDINLNTDNLTYKKS-- 66
DB 61 VEKIFYMHENQFKINDYNIILQYLIEYNNKICKENKFFCK--NPL---YNTYKCKLY 115

QY 67 --GRNKALKKPKVLTWEVLELKNLSLTPVEKNLHFIWIGQINDTAINYINQKDVNSD 124
DB 116 IYLDYBEKKDKELVIN--TEQKN-----AVDK-----INDKNVNIHSNETI 159

QY 125 YTVKVPYDSNAFLNTLTKTIVESATNTLESFRENLPDPEYKFKRMELIYDKQK 184
DB 160 ITGK--ETLIDILNKLK-LVSDKQLIEQYKNNKEIEFKN----- 201

QY 185 HFIDYKSOIEENPEFIDNIKTYSNEYSKOLEALNKYIBESLNKIKITANNQNDIRNLE 244
DB 202 -----IDNVQK-----EINKQDELNKLKLSKKEFFIKQ-----EELN 235

QY 245 KPADEDIVLYNQBLVERWNLAAASDILRISMLKEDGGVYLDVLDILPGIQDPLFKINP 304
DB 236 KTIDKK-----QEBELIKLN-----DKEINFNID-----EKQKLDQI 268

QY 305 DSINTSWMTKLEAIKMK-YKEYIPGYTSKFNDFMLDEEVQSPESALSSKSDSEIFLPL 363
DB 269 NSKINTLNENIK--GVNWLTE-----TKNKSINLQNEILNK-DSTIKSLDEKQLDEL 320

QY 364 DDIVKSPLEVKIAPANNVIN-QALI--SLKDSYCSDLVINQKNRYKI-----LNQN 413
DB 321 DK-NINNTISLYNKSNTKITNIQQLLESSLTDFNANINELKSKIKLFDNDIQKLNND 379

QY 414 LNPISNGTD-FNTTMKIPSDKLAS-----TSN-----EDNMPMKITNLYLVKQFA 459
DB 380 ITEQNKTIDPFNNSTRIFKEKLDTEYKIDDIKNNLQKLESYKKIDQETYYKKNKIN 439

QY 460 PDVASTINLSPGV-----YTGAQDLMLFKONSTNHLLEPRL 499
DB 440 KEYNDIIEELKNNLQKLEENKKIDQETYYKKNKINKEYNDIIEELKN--NLOKLEENK 497

QY 500 NPEFPKTKISLTQETITSLMSFNO-----ARAKSOFEPYKGYEGEGALGED 546
DB 498 NINDKLYKNDIESN-TEL--FNKLNISDFKDKSRIAKLNTYEYQLKOLLENKTN 554

QY 547 DNLDFQNTV--LDKDYVSK-----KILSSMKTRNKYIHYIVOLQGDKISYEASCN 596
DB 555 ELMKLSDNKLSLEQLYDSKKNLDGIDKIYNSLKEKN-----DKI----- 595

QY 597 LFSKDPYSSILYQKNIEGSETAYYYVADAEIKEDIKRYPYQISNKNKIKLFIHGKKS 656
DB 596 ----DEYFS-----NIEKFDYNTVENK-----FIGNLD 621

QY 657 ----EFNTDTF--ANLQDVSLSEIETILNADJSPKYIEINLIGQNMESYSISABE 709
DB 622 IINKINDQPKYINSKIDSKNELSTMFD-----DIFNAKNQIASINTNIENIS 672

QY 710 TYPGKLLLIKORVSELMPSISQDSITVSANQVEVRINERKGRBILDHSGKWINKERSII 769
DB 673 -----NKIKD-----LNERTIS-NEDSSKELLDEIRKY-KQQPDKI 706

QY 770 KDISKEYISNPENKII--VSKYLHELSTLLQBRNNANSSDIDLEKKVWLTECEI- 826
DB 707 KQAVNTEVKSFENTLQKDIDSISKN-INELTNAYDIINTKAN-----DLDDKLNYSGEFK 761

QY 827 --NVASNIDRQIVEGRIEBAKULT-----SDSINYIKNEF--KLIE-----SISDS 868
DB 762 NLYNASDOLLDTIQNNDEKVKQLNVEKKNQKQSIINDIVNFIKELIKFNNTENKS 821

QY 869 LYDLKHQGLDSDSHFISFEDISKTEGFRIRFINKETGN-----SIFETEKEIFS---- 919
DB 822 LNELTNDNDINDKIFKLYKELNKISTNNLLIKYKNEIDNVNKLISVIENTLQFINSLSI 881

QY 920 -----EVATHIS-----KEISNIKIDIFDNVNGKIVKYV 948
DB 882 EFNQGSITSHFNFLNTLAGINDVLNKLKIMADTTRRGDTNIRDKQISSENIKSQ 941

QY 949 NLDAAEVNTLNSAFFIQSLIEYNTT-----KESLSNLSVAMKVQVY 990
DB 942 KFNKNEKDLKLLISFNDKLNKYNISAGYTEYNNIEHECLKLYLIVSDQEY 993

RESULT 10
T18427
Hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C/Accession: T18427
R:Lawson, D.; Bowman, S.; Barrell, B.
Submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18427
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3724 <LAW>
A:Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325379; PIDN:CAB11104.1
C:Genetics:
A:Introns: 307/1; 1545/2
A:Note: C0335c

Query Match 6.3%; Score 328; DB 2; Length 3724;
Best Local Similarity 21.9%; Pred. No. 0.00014;
Matches 230; Conservative 173; Mismatches 347; Indels 302; Gaps 57;

QY 20 QDEYVAIINA-----LEEYHNMSESSVVEKYLKLDINLNTDNLTYKKS-----G 67
DB 618 EDEETIHLVKENLKDANEYNNDKEN-----KNWTKELK-SKYLENEKGTLELKLGR 672

QY 668 RNKALKKFKFYLTMEVLELKNLSLTPVEKNLHFIWIGQINDTAINYINQKDVNSDYTV 127
DB 673 KNNIFKDEKY-----NSLGEV-----IINEIQIENKINDIQDG-- 708

QY 128 KVFYDSNAFLINTLTKTIVESA--TNNTLESFRNLNDPEFYKFKRMELIYDKQKH 185
DB 709 -----NISKQKIIQSSRTNDFNFKDISLAND---DLEKERRK-----KSQH 748

QY 186 FIDYY-----KSQIENPEFIIDNIKTY-----LSNEXSKDL-----BALNKYIBESLNKI 232
DB 749 FIDNLVADKQNEISEINIKICDNNINNYDESINNIDESINNIDESINNIDESINNI 808

QY 233 TANNNDI--RNLEKFADEDLVRLYNQ-----ELVERWNLAASDILRISMLKEDGG 282
DB 809 YDENINNIYDENINNIYDENINNIYDEGINIKICDNNILENKNIKTNDIVQV---EENN 864

QY 283 VYLDVDVILPGIQDPLFKSINKPDSITNTSWEMIKLEALMKYKEIPIGVTSKNFMDLDEV 342
DB 865 ESIE-----KNELMISLNKD---INNTYNNFK-----ENVDIFINKI 898

QY 343 QRSFESALSSKSDKSEIFLDDIKVSPLEVKIAPANNVINQALISLKSQYCSDLVINQ 402


```

Qy      893  ENGPRIRFINKTGNSTFIETEKEIFSEYATHISKESINIKDTIFNVNGKLVKKVNLDA 952
          |||       :|||       :|||       :|||       :|||       :|||       :|||
Db     1014  INGKYNEVISNYRGYS-----ENISSKLENIEMHEIENLSRRLTDRIID--SLSKGMD--- 1063
          |||       :|||       :|||       :|||       :|||       :|||       :|||
Qy      953  AHEVNTLNSAFFIOSLIYEYNTTKESLNSIVAMKVQVY-AQLPSTGLNITTD--ASKAVE 1009
          |||       :|||       :|||       :|||       :|||       :|||       :|||
Db     1064  -----ENLOKLKESFDVSKYQVEFKELVKVDLTDDGEAKINK 1100
          |||       :|||       :|||       :|||       :|||       :|||       :|||
Qy     1010  LV-----STALDETID 1020
          |||       :|||       :|||       :|||       :|||       :|||       :|||
Db     1101  LVKEIEQYKKSRLUEAID 1118
          |||       :|||       :|||       :|||       :|||       :|||       :|||

RESULT 12
Ti18501
hypothetical protein C0760C - malaria parasite (Plasmodium falciparum)
C/Species: Plasmodium falciparum
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C/Accession: Ti18501
R/Lawson, D.; Bowman, S.; Barrrell, B.
submitted to the EMBL Data Library, August 1997
A/Reference number: Z18935
A/Accession: Ti18501
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-3394 <LAW>
A/Cross-references: EMBL:Z98551; NID:e1331903; PID:e1331922; PIDN:CAB11140.1
C/Genetics:
A/Map position: 3
A/Note: C0760C

```

[illegible]

Db	2236	KYIKIMNNIIEKDEIIEINLKNKYNKLDLNNYSVVDKSVCSFEDSNIMSPSCNDIL	22995
Qy	399	-VINQIKNRK-----ILNDLN--PSINEGTDFNTWKIPS-DKLASISNEDNNMFM	447
Db	2296	NVFNILSKSKVKCTMDICENWDSISINNVNINNVNINNVNINNVNINNVNINVKI	2355
Qy	448	IKITNYLKVGFADPVRSTNLSGPGVYTAYODLLMFKONSTNIHLLPELRFEPFKTK	507
Db	2356	VDINNYLVNKL-----QLNKONDNIIIKFNILKFLGSCVLYINRNKIEIQMLKNQ	2409
Qy	508	ISGLTQBITSLWSF--NQARAKSQEPEYKGYFEGALGEDDNLDAQNTYLD-KDYVSK	564
Db	2410	ILSL-ESIKSLNEFTNLLKNENKELIKNNFEEILKKNLQDNESICQNLNLYLK	2468
Qy	565	-KILSSMKTN---KEYTHIVQLQGDKISYEASCNLFSDKPSYILYQKNEG--SE	616
Db	2469	NEELINKVKNVIKYGYIHLIQ-----QSNVCKIKFHNENKIIQSLINKLLYLK	2522
Qy	617	TAYVYVADAEIKEDIRIPYQISNKRMIKLTFIGHKSFEFTDTFANLDVDSLSEIE	676
Db	2523	KSPDFWYDVGIE-----IREKNIII-----NQDPLTDEV-----FKHIQ	2559
Qy	677	TIINLAKADISPKYIHI-----NLIG-----CMPSYS-----	704
Db	2560	TFTKTCNVLIQGYLSILKDTNNDFFIQNKQSQGNQNGHINMCNIYDPDDEINVTADQ	2619
Qy	705	-----ISAEETYPGKLL-----	717
Db	2620	QIPDGTENVQOSIQNEEDYVNNEMVTDKMLDNNRNGDDDDDDDDDDNNNNNNNN	2679
Qy	718	-----KIKRVSELMPSISODSITVSANQVEVRINBEGKREIILDHSGKWINKESIIKDI	772
Db	2680	NNNNNMGDEDLHNAFNHLLTNGVKSQDINNTELRY-----EENIIQNI	2729
Qy	773	SSKEYISFN--PKENKIIVSKY-----LHELSTLQETRNANSSDIDLEKK	818
Db	2730	YTNDVNDNVNQVIENINKLIKDKQDIINDELKNEHNLIIRLINESTENAH-----NLENV	2785
Qy	819	VMLTECEINVASMIDQIVEGRETEAKULTSDS-----INVIKNEFKLIESI	865
Db	2786	YVQNDANNLINDIKKEETIYTYDEKDNVSNEGSKCDDDKENEDIIQAKNENPVSTH	2845
Qy	866	SDSLYDLKHONGDDSHFISDFISKTENGFRIRFINKETGNSIFITEKEIFSEYATHI	925
Db	2846	YDNDDINKDNINNDNNNDINDDNNDN-----INNDNNND-----NINNDINND	2892
Qy	926	SKEISNIKTQIFDNVNGKLV-----KKVNLDAAEHVNTLNSAFFIQSLIEYN	972
Db	2893	NINNDNNNDNNDSNNGFVCELSSNINDFNILNVKDNFQINKSN--FSTNLSEYN	2950
RESULT 13			
GB1702			
adherence factor TC0437 [imported] - Chlamydia muridarum (strain Nigg)			
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn			
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000			
C:Accession: GB1702			
C:Read, T.D.; Bruntan, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.;			
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.;			
Nucleic Acids Res. 28, 1397-1406, 2000			
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumonia			
A:Reference number: AB1500; MUID:20150255; PMID:10684935			
A:Accession: GB1702			
A:Molecule type: DNA			
A:Status: preliminary			
A:Residues: 1-3255 <P>			
A:Cross-references: GB:AE002311; GB:AE002160; NID:g7190477; PIDN:AAF39291.1; PI			
A:Experimental source: strain Nigg (MoPn)			
C:Genetics:			
A:Gene: TC0437			

Query Match 6.1%; Score 316.5; DB 2; Length 3255;

Db 1123 ELAK-KYLNEQRVHVHLEIDDYLTNPLFNRLHEEGVAFSDLTITRTWYLAEGYIS----- 1176

Qy 833 DROIVEGRIEEAKNLTSDSI-----NYIKNFKLIESIDSLYDLKHONG---LDDSH 882

Db 1177 -----GIFSEGNILPSPSARLVNIKITVYGGDYHDMQDVLPIXYDLASGGAADLTNER 1230

Qy 883 FIFPED-----ISKTENGFPIRFINKETGNSIFETETEKEIFS---EY 921

Db 1231 FAAPESLRKNVLEGLHGTDLTTPVDASVSGWGSF-GVENG-----VESDHTMISIAPGF 1285

Qy 922 ATHISKEISNIKTIFD-----NVNGKLVKKNVNLDAAEHVNTLNSAFFI-----QSL 968

Db 1286 PNGASYSQMYHLSALVEIHRHILGSLTSLDIKK-----ELESKAGCFVHERPDSL 1338

Qy 969 IEYNTTKE--SLNSLNVAMKVQVY-AQLFTGLNT-ITDASKVVE 1009

Db 1339 LKASSEQYLSLTETIHKSLSNQVHLAEAVSHLMTALPGVGKIIE 1383

RESULT 14

JC6009

surface-located membrane protein lmp3 precursor - Mycoplasma hominis

C:Species: Mycoplasma hominis

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C:Accession: JC6009

R:Ladefoged, S.A.; Jensen, L.T.; Brock, B.; Birkelund, S.; Christiansen, G.

J. Bacteriol. 178, 2775-2784, 1996

A:Title: Analysis of 0.5-kilobase-pair repeats in the Mycoplasma hominis lmp gene system

A:Reference number: JC6009; MUID:96213016; PMID:8631664

A:Accession: JC6009

A:Molecule type: DNA

A:Residues: 1-1302 <LAD>

A:Cross-references: EMBL:X95601; NID:G1197335; PIDN:CAA64858.1; PID:G1197336

C:Genetics:

A:Gene: lmp3

A:Genetic code: SGC3

C:Superfamily: surface-located membrane protein lmp3; tetratricopeptide repeat homolog.

C:Keywords: duplication; membrane protein

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-1302/Product: surface-located membrane protein lmp3 #status predicted <MAT>

F:957-992/Domain: tetratricopeptide repeat homology <TT1>

F:993-1026/Domain: tetratricopeptide repeat homology <TT2>

F:1089-1120/Domain: tetratricopeptide repeat homology <TT3>

F:1154-1190/Domain: tetratricopeptide repeat homology <TT4>

Query Match 6.1%; Score 315; DB 1; Length 1302;

Best Local Similarity 20.2%; Pred. No. 0.00011;

Matches 235; Conservative 203; Mismatches 396; Indels 330; Gaps 53;

Qy 3 LVNKAQIQKMYVK-FRIQDEYVAILNALAEYH-----NMSESSVVE--- 44

Db 169 LNKINAERELQSIKFNEKQELKRVLD-LEDTKVEDFTKEQKVFETININETSIEDIK 227

Qy 45 -KYLK-KDINLTDNVINYKSGRNKALKKFKYELTMEVLKKNLSLTPEKNLHFIW 102

Db 228 NKIIIEVKATSSLTSLKLT-----KQELQEF-ENIKXQLQDFINTKLNDAKYO----- 276

Qy 103 IGGQINDTAINYNOWKDVNSDYTVKVFYDSNAFLNTLTKKTVESAT---NNTLES PRE 159

Db 277 ---SIKQALDKXNSLNGINKNSTIKI KAGQNALIKAKEEAGLEKEKLDQNLKDTLKE 333

Qy 160 NLADPEPDYKFKYRKRMEIIVDKHFIDYKQSIQIENPEFIIDNIKTYLSNYSKDL 219

Db 334 TINNAK-----BPKLLIDNDQKIVD-LKSNLDNEISKAQ 368

Qy 220 ALNKYIEESLNKITTANNNDIRNLEKFADEDVLVYNQELVERWNLAASDILISMKE 279

Db 369 SLSK-----DKESMESANDLLNKLIEYKELNKNFQKEAKFN-----ELEQTRK 414

Qy 280 DGGVYLDVDI-----LPQIQDLFKSINKPDSINTSWEMIKLEAIMKYEIVPGYTSKN 334

Db 415 NIENFLTDEVKNNPNYATLVKDLTNADKAKSVINSS---NKSDIIAANEALIQALADAN 471


```

QY 776 EYISPNKEN-----KIIVSK-YLHELSTLLQEIERNAN 809
: : : : :
Db 836 LQWVSESQNHSDINQCLNEVANIYILKLNKIKKIIDKKEYTSEIEKNKNQINDELN 895
: : : : :
QY 810 SSDIDLEK---KWLTECEINVASNIDROIVEGRIB-----AKNLTSDSINYIKNEFK-- 860
: : : : :
Db 896 NSEKVIKIEGDLSEKCRSKINSTLDDKDIDECIKNINVLKKNILNEBTN-ITNHFKN 954
: : : : :
QY 861 -----LIESISDLYDLKHONGLDDSHFISPEDISK---TENGPRIRFIN 902
: : : : :
Db 955 EYINKIVLSNFNNIEMADNKSQVILEIKQNGTND-HDYNIKELKSHKXNGYK---T 1009
: : : : :
QY 903 KETGNSIFTETEKEIPSEYATHI-----SKEISNIKDTIFDNVNGKLVKKNLDAAH 954
: : : : :
Db 1010 EADONKKAIOKNKELFEQYKEVTVLLNKYVAVELKNKEDKT-KNDSKQIIEIK--DAH 1066
: : : : :
QY 955 EVNTLNSAFPQISL-----IEYNTTKESLSNLSVAMKVQVYAQLFSTGLNITD 1003
: : : : :
Db 1067 NYCTLESQKSEKQWNEIKNEKIHIEDEVANNDKSNKAIT-SIKVSEPFKTIKINE 1123
: : : : :

```

Search completed! April 1, 2004, 16:45:02
Job time : 22.25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 16:44:46 ; Search time 34.125 Seconds
(without alignments)
7840.786 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1020
Perfect score: 5189
Sequence: 1 MNLVNRQAQLQKMYVFRQ.....ITDASKVELVSTALDETID 1020

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1069545 seqs, 262320428 residues

Total number of hits satisfying chosen parameters: 1069545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4101	79.0	2366	14	US-10-011-366-10
2	4101	79.0	2366	15	US-10-354-774-10
3	4101	79.0	2366	15	US-10-271-012-10
4	2791	53.8	2710	14	US-10-011-366-6
5	2791	53.8	2710	15	US-10-354-774-6
6	2791	53.8	2710	15	US-10-271-012-6
7	2204	42.5	556	12	US-10-463-957-1
8	2193	42.3	556	12	US-10-463-957-11
9	2190	42.2	556	12	US-10-463-957-12
10	2051.5	39.5	522	12	US-10-463-957-9
11	1979	38.1	500	12	US-10-463-957-3
12	1734	33.4	420	12	US-10-463-957-5
13	769	14.8	170	12	US-10-463-957-7
14	422.5	8.1	3169	14	US-10-114-170-257
15	290.5	5.6	4688	12	US-10-282-122A-76865

```

16 279 5.4 1178 12 US-10-282-122A-52434 Sequence 52434, A
17 276.5 5.3 1182 12 US-10-282-122A-53445 Sequence 53445, A
18 275.5 5.3 1279 12 US-10-282-122A-52455 Sequence 52455, A
19 270 5.2 1196 12 US-10-282-122A-52737 Sequence 52737, A
20 265 5.1 4620 12 US-10-282-122A-68921 Sequence 68921, A
21 257.5 5.0 1090 15 US-10-369-493-18439 Sequence 18439, A
22 257.5 5.0 1162 15 US-10-452-024-113 Sequence 113, App
23 257.5 5.0 1162 15 US-10-452-024-114 Sequence 114, App
24 257.5 5.0 2184 14 US-10-304-095-6 Sequence 6, Appli
25 256 4.9 903 12 US-10-282-122A-5328 Sequence 5328, A
26 255 4.9 1875 15 US-10-369-493-22285 Sequence 22285, A
27 253.5 4.9 1184 12 US-10-282-122A-53254 Sequence 53254, A
28 252 4.9 1948 14 US-10-032-585-7611 Sequence 7611, Ap
29 249 4.8 1679 15 US-10-369-493-22080 Sequence 22080, A
30 245.5 4.8 1275 12 US-10-282-122A-53272 Sequence 53272, A
31 245.5 4.7 1103 12 US-10-282-122A-76866 Sequence 76866, A
32 245.5 4.7 1881 14 US-10-032-585-7646 Sequence 7646, Ap
33 245 4.7 1639 14 US-10-087-464-10 Sequence 10, Appl
34 244.5 4.7 6641 12 US-10-282-122A-70580 Sequence 70580, A
35 244 4.7 1143 15 US-10-369-493-11081 Sequence 11081, A
36 243 4.7 1928 15 US-10-369-493-22025 Sequence 22025, A
37 242 4.7 1005 15 US-10-369-493-1061 Sequence 1061, Ap
38 241 4.6 1494 12 US-10-282-122A-47128 Sequence 47128, A
39 239.5 4.6 1957 15 US-10-369-493-2070 Sequence 2070, Ap
40 239 4.6 1837 15 US-10-369-493-22734 Sequence 22734, A
41 237.5 4.6 1162 15 US-10-452-024-115 Sequence 115, App
42 237.5 4.6 5005 12 US-10-282-122A-76871 Sequence 76871, A
43 237 4.6 1163 12 US-10-282-122A-51864 Sequence 51864, A
44 236.5 4.6 1790 15 US-10-369-493-1586 Sequence 1586, Ap
45 235.5 4.5 1155 12 US-10-282-122A-53419 Sequence 53419, A

```

ALIGNMENTS

RESULT 1

US-10-011-366-10
; Sequence 10, Application US/10011366
; Publication No. US20030054493A1

GENERAL INFORMATION:

APPLICANT: Williams, James A.

APPLICANT: Kink, John A.

TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES

OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE

DISEASE

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Medlen & Carroll

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/011.366

FILING DATE: 16-NO. US20030054493A1-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/957.310

FILING DATE: 23-OCT-1997

APPLICATION NUMBER: US 08/329.154

FILING DATE: 24-OCT-1994

APPLICATION NUMBER: US 08/161.907

FILING DATE: 02-DEC-1993

APPLICATION NUMBER: US 07/985.321

FILING DATE: 04-DEC-1992

APPLICATION NUMBER: US 07/429.791

FILING DATE: 31-OCT-1989

[illegible]

```

Qy 121 VNSDYTVKVFYDSNAFLINTLTKTIVESAANNLTLESFRENLDNPEFDYKFKRMEIY 180
Db 121 VNSDYNVNFYDSNAFLINTLTKTIVESAANNLTLESFRENLDNPEFDYKFKRMEIY 180
Qy 181 DKQKHFIDYKSKQIEENPEIINIKTYLSNEYSKDEALNKYIESLNKINTANGNDI 240
Db 181 DKQKHFIDYKSKQIEENPEIINIKTYLSNEYSKDEALNKYIESLNKINTANGNDI 240
Qy 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILIRISMLKEDGGVYLDVILPGIQDLPFS 300
Db 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILIRISMLKEDGGVYLDVILPGIQDLPFS 300
Qy 301 INKPSITNTSWEMIKLEAIKMYKEYIPGYTSKNFDMLEDEVOQSFESALSSKSKSEIF 360
Db 301 INKPSITNTSWEMIKLEAIKMYKEYIPGYTSKNFDMLEDEVOQSFESALSSKSKSEIF 360
Qy 361 LPLDDIKVSPLEVKIAFANNVINOALISLKDYSQDLVINQIKRYKILNDNLNPSINE 420
Db 361 LPLDDIKVSPLEVKIAFANNVINOALISLKDYSQDLVINQIKRYKILNDNLNPSINE 420
Qy 421 GTDFTNTMTKIFSDKLASINEDNMFMKIKITNYLVKGFAPDVRSNTINLSGPGVYTCAYOD 480
Db 421 GTDFTNTMTKIFSDKLASINEDNMFMKIKITNYLVKGFAPDVRSNTINLSGPGVYTCAYOD 480
Qy 481 LLMFKGDSNTHLLPELRNPFPPKTKISOLTEQITSLWSFNQARAKSQFEYKGYE 540
Db 481 LLMFKGDSNTHLLPELRNPFPPKTKISOLTEQITSLWSFNQARAKSQFEYKGYE 540
Qy 541 GALGEDNLDPAQNTVDKQYVSKKILSMKTRKNEYIHYIVQLQDKISYEASCNLFK 600
Db 541 GALGEDNLDPAQNTVDKQYVSKKILSMKTRKNEYIHYIVQLQDKISYEASCNLFK 600
Qy 601 DPYSILYQKNIQESGTAIYVYVADAIKIDKIRIPYQISNKNKILTIHGKGEFNT 660
Db 601 DPYSILYQKNIQESGTAIYVYVADAIKIDKIRIPYQISNKNKILTIHGKGEFNT 660
Qy 661 DTFANLDVSSSIEITNLAKADISPKYIEINLLGNMFSYSISAEETYPCKLLKIK 720
Db 661 DTFANLDVSSSIEITNLAKADISPKYIEINLLGNMFSYSISAEETYPCKLLKIK 720
Qy 721 DRSELMPISQDSITVSANQYVIRNEEGKREILDHSGKWINKESIIKDISKEYISF 780
Db 721 DRSELMPISQDSITVSANQYVIRNEEGKREILDHSGKWINKESIIKDISKEYISF 780
Qy 781 NPENKIIVKSKYLHEISSLQELIRNNANSDDILEKKVMLTECEINVASNIDRQIVEGR 840
Db 781 NPENKIIVKSKYLHEISSLQELIRNNANSDDILEKKVMLTECEINVASNIDRQIVEGR 840
Qy 841 IEEAKNLTSDSINVIKNEFKLIESISDLYDLKHONGLDSDSHFISFEDISKTENGPRIF 900
Db 841 IEEAKNLTSDSINVIKNEFKLIESISDLYDLKHONGLDSDSHFISFEDISKTENGPRIF 900
Qy 901 INKGTGNSIFTETKEIFSEVATHISKEISNIKDTIFDNVNGKLVKKNVLDAAHEVNTLN 960
Db 901 INKGTGNSIFTETKEIFSEVATHISKEISNIKDTIFDNVNGKLVKKNVLDAAHEVNTLN 960
Qy 961 SAFTIQSLIEYNTKESISNLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
Db 961 SAFTIQSLIEYNTKESISNLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020

```

RESULT 3
 US-10-271-012-10
 ; Sequence 10, Application US/10271012
 ; Publication No. US20030219457A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Williams, James A.
 ; Thalley, Bruce S.
 ; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
 ; Botulinum Neurotoxin
 ; NUMBER OF SEQUENCES: 82
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Medlien & Carroll

```

; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/271,012
; FILING DATE: 15-Oct-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-271-012-10
Query Match 79.0%; Score 4101; DB 15; Length 2366;
Best Local Similarity 77.6%; Pred. No. 4.8e-250;
Matches 792; Conservative 106; Mismatches 122; Indels 0; Gaps 0;
Qy 1 MNLVAKAQLQKVVYKFERIQDEYVAIILNALLEEYHNHSESSWEKYLKLDINLNTDYL 60
Db 1 MSLVNRKQLERKMANVFRTOQEYVAIILDALEEYHNHSESSWEKYLKLDINLNTDYL 60
Qy 61 NTKKSGRNKALKKPKKYLTVMEVLKNSLTPVEKNLHFIWIGQINDTAINYNQWD 120
Db 61 DTKKSGRNKALKKPKKYLTVMEVLKNNLTPVEKNLHFIWIGQINDTAINYNQWD 120
Qy 121 VNSDYTVKVFYDSNAFLINTLTKTIVESAANNLTLESFRENLDNPEFDYKFKRMEIY 180
Db 121 VNSDYNVNFYDSNAFLINTLTKTIVESAANNLTLESFRENLDNPEFDYKFKRMEIY 180
Qy 181 DKQKHFIDYKSKQIEENPEIINIKTYLSNEYSKDEALNKYIESLNKINTANGNDI 240
Db 181 DKQKHFIDYKSKQIEENPEIINIKTYLSNEYSKDEALNKYIESLNKINTANGNDI 240
Qy 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILIRISMLKEDGGVYLDVILPGIQDLPFS 300
Db 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILIRISMLKEDGGVYLDVILPGIQDLPFS 300
Qy 301 INKPSITNTSWEMIKLEAIKMYKEYIPGYTSKNFDMLEDEVOQSFESALSSKSKSEIF 360
Db 301 IEKPSSTVDVFWMTKLEAIKMYKEYIPGYTSEHFDMLDEVOQSFESALSSKSKSEIF 360
Qy 361 LPLDDIKVSPLEVKIAFANNVINOALISLKDYSQDLVINQIKRYKILNDNLNPSINE 420
Db 361 LPLDDIKVSPLEVKIAFANNVINOALISLKDYSQDLVINQIKRYKILNDNLNPSINE 420
Qy 421 GTDFTNTMTKIFSDKLASINEDNMFMKIKITNYLVKGFAPDVRSNTINLSGPGVYTCAYOD 480
Db 421 DNDFTNTMTKIFSDKLASINEDNMFMKIKITNYLVKGFAPDVRSNTINLSGPGVYTCAYOD 480
Qy 481 LLMFKGDSNTHLLPELRNPFPPKTKISOLTEQITSLWSFNQARAKSQFEYKGYE 540
Db 481 LLMFKGDSNTHLLPELRNPFPPKTKISOLTEQITSLWSFNQARAKSQFEYKGYE 540

```

541 GALGEDNDLFAQNTVLDKDYVSKILSSMKTRNKYIHYIVQLQGDKISYEASCNLFK 600
 541 GSLGEDNDLFSQNVVDKEYLLEKISLARSSRGYIHYIVQLQGDKISYEACNLFAK 600
 601 DPYSILYQKNIQESGATAYVYVADAETDKYRIPYQISNKENIKLTIGHGKSEFNT 660
 601 TPYDSVLFPQKNIQESGATAYVYVADAETDKYRIPYQISNKENIKLTIGHGKSEFNT 660
 661 DTFANLDVSLSSIEITILNAKADISPKYIEINLLGNMFSYSISAEETYPGKLLIK 720
 661 DIFAGFDVSLSTIEAIDAIDAKEDISPKYIEINLLGNMFSYSINVEETYPGKLLIKV 720
 721 DRVSELMPSISQDSITVSANOYEVRIINEEGREILDHSGKWINKESIIKDISKEYISF 780
 721 DKISELMPSISQDSITVSANOYEVRIINEEGREILDHSGKWINKESIIKDISKEYISF 780
 781 NPKENKIVKYLHELSTILOETRNANSSDIDLEKKWMLTECEINVASNIDQIVEGR 840
 781 NPKENKIVKYLHELSTILOETRNANSSDIDLEKKWMLTECEINVASNIDQIVEGR 840
 841 TEEAKNLSDSINYIKNPKFLESISDLYDLKHONGLDSDHFTSFEDISKTENGRPIRF 900
 841 TEEAKNLSDSINYIKNPKFLESISDLYDLKHONGLDSDHFTSFEDISKTENGRPIRF 900
 901 INKETSIGFIEETEKEIEPSEYATHISKEISNIKDTIFDNVNGKLVKKVNLDAHEVNTLN 960
 901 INKETSIGFIEETEKEIEPSEYATHISKEISNIKDTIFDNVNGKLVKKVNLDAHEVNTLN 960
 961 SAFFQSIQIEYNTTKESISNLSVAMKVQVYQALFSTGLNTITDASKVVELVSTALDSTID 1020
 961 AAFQSIQIEYNTTKESISNLSVAMKVQVYQALFSTGLNTITDASKVVELVSTALDSTID 1020

RESULT 4

US-10-011-366-6
 Sequence 6, Application US/10011366
 Publication No. US20030054493A1
 GENERAL INFORMATION:
 APPLICANT: Williams, James A.
 Kink, John A.
 TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
 OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
 DISEASE
 NUMBER OF SEQUENCES: 22
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Medlen & Carroll
 STREET: 220 Montgomery Street, Suite 2200
 CITY: San Francisco
 STATE: California
 COUNTRY: United States of America
 ZIP: 94104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/011,366
 FILING DATE: 16-NOV-2003
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/957,310
 FILING DATE: 23-OCT-1997
 APPLICATION NUMBER: US/08/329,154
 FILING DATE: 24-OCT-1994
 APPLICATION NUMBER: US/08/161,907
 FILING DATE: 02-DEC-1993
 APPLICATION NUMBER: US/07/985,321
 FILING DATE: 04-DEC-1992
 APPLICATION NUMBER: US/07/429,791
 FILING DATE: 31-OCT-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Ingolia, Diane E.

REGISTRATION NUMBER: 40,027
 REFERENCE/DOCKET NUMBER: OPED-01121
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 705-8410
 TELEFAX: (415) 397-8338
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2710 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 6:
 US-10-011-366-6

Query Match 53.8%; Score 2791; DB 14; Length 2710;
 Best Local Similarity 52.1%; Pred. No. 3.6e-167;
 Matches 534; Conservative 221; Mismatches 263; Indels 8; Gaps 5;
 1 MNLVNAQLQKQVYKVFQIODEYVAIALAEVYHNMSESSVVEKYLKLDINLTNYL 60
 1 MSLISKEELIKLAY-SIRPRENEYKLTILNDEYNNKLTNNENKYLQKLINESIDVFM 59
 61 NNYKSGRNKALKKPEVLTWEVLELKNLSLTPVEKNLHFTWIGQINDTAINYNQMD 120
 60 NKYKTSNRNALSNLKKDILKEVILIKNSNTSPVEKNLHFTWIGVSDIALEYIKQAW 119
 121 VNSDVTWKVYDSNAFLNTLTKKTIVESATNTLSPFRENLMDFDYNKFKYRMEIY 180
 120 INAEYNKLVYSEAFVNTLKAIVESTTEALQLLEEIQNFQDNKMKYKKEMEIY 179
 181 DKQKIFIDYKQIEENPEFIIDNIIKTYLSNEYSKDLKALNKYTEESINKITANGNDI 240
 180 DRQKRPINYYKQINKPTVTIDDIKSHLVSEYNRDETIVLESYRTNSLRKINSNHGDI 239
 241 RNLEKPADEDVRLVNOELVERWNLAASDILIRISMLKEDGVYLDVLDLQIQDLEKS 300
 240 RANSLFTEQELLNIYSQELLNGLNLAASDILIRISMLKEDGVYLDVLDLQIQDLEKS 299
 301 INKPSDITNTSWEMIKLEAIMKYKEVIGYTSKFNFDMLDEEVORFESALSKSKSEIF 360
 300 ISRPSIGLDREMIKLEAIMKYKEVIGYTSKFNFDMLDEEVORFESALSKSKSEIF 359
 361 LPDDIKVSPLEVKYAFANNVINQALISLQSDYCSLDVINQIKRYKILNDNLNPSINE 420
 360 SKLENLVSDLEIKAFALGVSINQALISLQSDYCSLDVINQIKRYKILNDNLNPSINE 419
 421 GTDNFTMKIFSDKLASISNEDNMFMKITNYLVKGFAPDVRSTINISGPGVITGQD 480
 420 DNNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLVQGFMPPEARSTISLGGPAYSAYD 479
 481 LLMFKDNSTNIHLLEPELNPFPKTSQLEQITSLWFSNQARAKSQREYKGYFE 540
 480 FINLQENTIEKTKASDLIEFKPEPENNLSQLEQITSLWFSNQARAKSQREYKGYFE 539
 541 GALGEDNDLFAQNTVLDKDY-VSKKILSS--MKTRNKYIHYIVQLQGDKISYEASCNL 597
 540 GSLGEDNDLFSQNVVDKEYLLEKISLARSSRGYIHYIVQLQGDKISYEACNL 599
 598 FSXDPYSSILYQKNIQESGATAYVYVADAETDKYRIPYQISNKENIKLTIGHGK 655
 600 TPYDSVLFPQKNIQESGATAYVYVADAETDKYRIPYQISNKENIKLTIGHGK 657
 656 SEFNTTFTFANLDVSLSSIEITILNAKADISPKYIEINLLGNMFSYSISAEETYPGKL 715
 658 DEFNTSEFARLSVDSLNSISGFLDTIKLIDISPKVNEVNLGNMFSYDFNVEETYPGKL 717
 716 LKIKDRVSELMPISQDSITVSANOYEVRIINEEGREILDHSGKWINKESIIKDISK 775
 718 LLSIMDKITSTLPVKNKISITGANOEYRINSEGRKELLASHGKWINKESIIKDISK 777
 776 EYISFNPKNENKIIVKSKYLHELSTILOETRNANSSDIDLEKKWMLTECEINVASNIDQ 835
 778 EYIFFDSIDNKLKAKSKNIPGLASISEDITLLDASVSPDKFLLNKLKLINESSIGDY 837

Thalley, Bruce S.
TITLE OF INVENTION: Multivalent Vaccine For Clostridium Botulinum Neurotoxin
NUMBER OF SEQUENCES: 82
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/271.012
FILING DATE: 15-Oct-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/704.159
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPD-02304
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-271-012-6

Query Match 53.8%; Score 2791; DB 15; Length 2710;
Best Local Similarity 52.1%; Pred. No. 3.6e-167;
Matches 534; Conservative 221; Mismatches 262; Indels 8; Gaps 5;
QY 1 MNLVNAQLOKQVYKFIQDEVAIINLAEYHNMSESVVEKYLKLDINLTDNYL 60
DB 1 MSLISKEELIKLAY-SIRPRENEYKTIITLNDYENKLTNNENKYLQKLKINESIDVFM 59
QY 61 NTKYSGENKALKKFEVLTWVLEKNSLTPVEKXHLFWIGQINDTAINYQW 120
DB 60 NKYTSRNRALSNLKDDILKEVILIRKNSNTPVEKNLHFVWIGVSDIALEVIKOWAD 119
QY 121 VNSDYTVKVFYDSNAFLNTLTKTIVESATNNTLESFRENLDPEFDYKFKRMEIY 180
DB 120 INAEYIKWDSFAFLVTLTKALVESSTTEALQLLEEIQNFQDNMKEYKGEFIY 179
QY 181 DKQHFIDYKSQIEENPEFIDNIITKYLSEYKOLAEALNKYIEBSLNKITANNNDI 240
DB 180 DRQKFEINYKSIQKVPVTIDDIKSHLVSEYNRDETIVLESYRTNSLRKINSNHGIDI 239
QY 241 RNLEFAEDLVRLNQELVERAWNLAASDILRISMLKEDGGVYLDVILPGIOPDLFS 300
DB 240 RANSLFTEQELNITYSQELNARGNLAASDILRILALKNFGVYLDVDMLEPGHSDLFKT 299
QY 301 INKPDSTIITWEMTKLAIMKYKVIPIGYTSKNFDMLEDEEVQSFESALSKDKKSEIF 360
DB 300 ISRPSIGLDRWEMTKLAIMKYKVIINNYTSNFDKLDQQLKFNKLIIESKESEIF 359
QY 361 LPLDDIKVSPLEVKIAFANNVINOALISLKDYSCLVINGIKRYKILNDNLNPSINE 420
DB 360 SKLENLNSDLKIAFALGVSINOALISKQSGLTNLVIEQVKRYQFLNCHLNPALES 419
QY 421 GTDFNTMTKIFSDKXASINENNMFMKITYNLKVGFPADVRSTINISGPGVYTGAYQD 480

DB 420 DNNFTDTTKIFHDSLFNSATAENSMFLTKIAPYLOQVFMPEARSTISLSPGAYASAYD 479
QY 481 LLMPKDNSTNIHLEPELRNPEFPKTKISQLTQEBITSLMSFNQARAKSOFKYEYKGYE 540
DB 480 FINLQENTIEKTLKASDLIEFKFPENNLSQLTEQINSLSWFDQASAKYOFKYEYVDYTG 539
QY 541 GALGEDNLPFAQNTVLDKDY-VSKKILSS--WKTANKEVIHYIOLQGDKISYEASCNL 597
DB 540 GSLSEGDVDFNKVNTALDKNYLLNNKIPSNNEEAGSKNIVHYIIQGGDDISYEATCNL 599
QY 598 FSKDPYSSILYOKNIEGSETAYVYVAD--AEIKEIDKYRIPYQISNKRNIKLTIFIGHGK 655
DB 600 FSKNPKNSIIQRNN--NESAKSYFLSDDGESILELNKYZIPELKNKVEKVKVTFIGHGK 657
QY 656 SEFNTDTFANLVDVDSLSSEIETILNLAKADISPKYIENILLGNMFSYSISAEETYPGKL 715
DB 658 DEFNTSEFARLSVDSLSEI--SSFLDTIKLDISPKQVEVNLGCMFSDYDFNVEETYPGKL 717
QY 716 LLKTKDRVSELMPSISQDSITVSANQVEVINBEGKEIILDHSGKWKINKESIIDKISSK 775
DB 718 LLSIMDKITSLPDVNKNISITIGANQYEVINSEGRKELLASHGKWKINKBEAIMSDUSSK 777
QY 776 EYISFNPKENKIIVKSKYLHELSTLLOSIIRNNANSSDIDLEKKVMTCEINVASNIDRQ 835
DB 778 EYIFPDSIDNKLKAKSNIPGLASISIDIKTLILDASVSPDTKFIKLNKLNIESSIGDY 837
QY 836 IVEGRIEEAKNLTGDSINYIKNEFKLIESDSIYDLKHQNGLDSDHFIQFEDISKTEG 895
DB 838 IYVEKLEPVKNIHNSDDLLDEFNLJENVSDELYELKCLNLDKYLISFEDISKKNST 897
QY 896 FRIRPINKETGNSIFETEKEIFSEYATHYSKETSNIKDTIFDNNVGNLKVKNLDAHE 955
DB 898 YSVRFINKNGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVGNNLLDNIQLDHTSQ 957
QY 956 VNTLSAFPIOSLIEYNTKESLNSLVAMKVQVYQALFSTGLANTITDASKVVELVSTAL 1015
DB 958 VNTLNAAPFIQSLIDYSSNKVDLNDLSTSVKQVYQALFSTGLNTIYDSIOLVNLISNAV 1017
QY 1016 DETID 1020
DB 1018 NDTIN 1022

RESULT 7
US-10-463-957-1
; Sequence 1, Application US/10463957
; Publication NO. US20040028705A1
; GENERAL INFORMATION:
; APPLICANT: Ballard, Jimmy D.
; APPLICANT: Spvres, Lea M.
; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637
; CURRENT APPLICATION NUMBER: US/10/463,957
; CURRENT FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: 60/389,685
; PRIOR FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 556
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-10-463-957-1

Query Match 42.5%; Score 2204; DB 12; Length 556;
Best Local Similarity 75.0%; Pred. No. 5e-131;
Matches 417; Conservative 66; Mismatches 73; Indels 0; Gaps 0;
QY 1 MNLVNAQLOKQVYKFIQDEVAIINLAEYHNMSESVVEKYLKLDINLTDNYL 60
DB 1 MSLVNRKQLEKMANVFRFTQDEYVAILDALBEYHNMSENVTVEKYLKLDINSLTDIYI 60
QY 61 NTKYSGRNLKALKKFEVLTWVLEKNSLTPVEKXHLFWIGQINDTAINYQW 120

Db 61 DTYKSGRNKALKKFKKEYLVTEVLEKNNLTPVEKNLHFVWIGQINDTAINYNQWKD 120
QY 121 VNSDYTVKVPYDSNAFLINTLTKTIVESATNTLTLESFRENLDNPFDPYKFKRMEIYY 180
Db 121 VNSDYVNVFYDSNAFLINTLTKTVVESAINDTLESFRENLDNPFDPYKFKRMEIYY 180
QY 181 DKQKHFDYKSOIENPEFIIDNIKTYSNEYSKOLEALNKYIEESLNKITANGNDI 240
Db 181 DKQKNFYNYKQREENPELIIDDIKTVLSNEYSKEIDELNTYIEESLNKITQNSGNDV 240
QY 241 RNLEKFADELDVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVILPGIQDPLFKS 300
Db 241 RNFEFKNGESFNLYEQELVERWNLAAASDILRISALKETGGMYLDVDMPLFGIQDPLFES 300
QY 301 INKPDSTNTSWEMIKLEAIMKYEIPGYTSKNFDMLEDEEVORSFESALSSKSKSEIF 360
Db 301 IEKPSVTVDFWMTKLEAIMKYEIPETYSHEFDMLEDEEVORSFESVLSKSKSEIF 360
QY 361 LPDDIKVSPLEVKIAFANNVINOALISLSDSYCSDLVINQIKRYKILNDMLNPSINE 420
Db 361 SSLGDMEASPLEVKIAFNKGIINOGLISVKDSYXSNLIVKQIENRYKILNNSLNPASE 420
QY 421 GTDFNTMTKIFSDKLASINEDNMFMKITNYLVKGFAPDVRSTINLSGGVYTCAYQD 480
Db 421 DNDFTNTTNTFIDSIMAEANADNGRFMMELGKYLVRGFPDVKTTINLSGPEYAAAYQD 480
QY 481 LLMFKDNSTNIHLEPELRNFEPPKTKISQLTEQITSLWSFNQARAKSQFEEYKKGYPE 540
Db 481 LLMFKESGMNIHLEADLRNFEISKTNISQTSQEMASLWSFDDARAKAQFEEYKKNRYFE 540
QY 541 GALGEDNDLDPQNTV 556
Db 541 GSLGEDNDLDFSQNIV 556

RESULT 8
US-10-463-957-11
; Sequence 11, Application US/10463957
; Publication No. US20040028705A1
; GENERAL INFORMATION:
; APPLICANT: Ballard, Jimmy D.
; APPLICANT: Spvres, Lea M.
; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637
; CURRENT APPLICATION NUMBER: US/10/463,957
; CURRENT FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: 60/389,685
; PRIOR FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 11
; LENGTH: 556
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Completely synthesized
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (395)..(395)
; OTHER INFORMATION: Xaa at position 395 is ala, asp, glu, phe, gly, his, ile, lys,
; OTHER INFORMATION: leu, met, asn, pro, gln, arg, ser, thr, val, trp, or tyr.
US-10-463-957-11

Query Match 42.3%; Score 2193; DB 12; Length 556;
Best Local Similarity 74.8%; Pred. No. 2.5e-130;
Matches 416; Conservative 66; Mismatches 74; Indels 0; Gaps 0;
QY 1 MNLVNAQLOKQVYKFIQDEYVAILNLEEEYHNMSSESVVEKYLKLDIINLTNYL 60
Db 1 MSLVNRKQLEKXANVRFTQDEYVAILDALEEEYHNMSSENTVVEKYLKLDIINLTDIYI 60
QY 61 NTKYKSGRNKALKKFKKEYLVTEVLEKNNLTPVEKNLHFVWIGQINDTAINYNQWKD 120

Db 61 DTYKSGRNKALKKFKKEYLVTEVLEKNNLTPVEKNLHFVWIGQINDTAINYNQWKD 120
QY 121 VNSDYTVKVPYDSNAFLINTLTKTIVESATNTLTLESFRENLDNPFDPYKFKRMEIYY 180
Db 121 VNSDYVNVFYDSNAFLINTLTKTVVESAINDTLESFRENLDNPFDPYKFKRMEIYY 180
QY 181 DKQKHFDYKSOIENPEFIIDNIKTYSNEYSKOLEALNKYIEESLNKITANGNDI 240
Db 181 DKQKNFYNYKQREENPELIIDDIKTVLSNEYSKEIDELNTYIEESLNKITQNSGNDV 240
QY 241 RNLEKFADELDVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVILPGIQDPLFKS 300
Db 241 RNFEFKNGESFNLYEQELVERWNLAAASDILRISALKETGGMYLDVDMPLFGIQDPLFES 300
QY 301 INKPDSTNTSWEMIKLEAIMKYEIPGYTSKNFDMLEDEEVORSFESALSSKSKSEIF 360
Db 301 IEKPSVTVDFWMTKLEAIMKYEIPETYSHEFDMLEDEEVORSFESVLSKSKSEIF 360
QY 361 LPDDIKVSPLEVKIAFANNVINOALISLSDSYCSDLVINQIKRYKILNDMLNPSINE 420
Db 361 SSLGDMEASPLEVKIAFNKGIINOGLISVKDSYXSNLIVKQIENRYKILNNSLNPASE 420
QY 421 GTDFNTMTKIFSDKLASINEDNMFMKITNYLVKGFAPDVRSTINLSGGVYTCAYQD 480
Db 421 DNDFTNTTNTFIDSIMAEANADNGRFMMELGKYLVRGFPDVKTTINLSGPEYAAAYQD 480
QY 481 LLMFKDNSTNIHLEPELRNFEPPKTKISQLTEQITSLWSFNQARAKSQFEEYKKGYPE 540
Db 481 LLMFKESGMNIHLEADLRNFEISKTNISQTSQEMASLWSFDDARAKAQFEEYKKNRYFE 540
QY 541 GALGEDNDLDPQNTV 556
Db 541 GSLGEDNDLDFSQNIV 556

RESULT 9
US-10-463-957-12
; Sequence 12, Application US/10463957
; Publication No. US20040028705A1
; GENERAL INFORMATION:
; APPLICANT: Ballard, Jimmy D.
; APPLICANT: Spvres, Lea M.
; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637
; CURRENT APPLICATION NUMBER: US/10/463,957
; CURRENT FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: 60/389,685
; PRIOR FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 12
; LENGTH: 556
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-10-463-957-12

Query Match 42.2%; Score 2190; DB 12; Length 556;
Best Local Similarity 74.8%; Pred. No. 3.8e-130;
Matches 416; Conservative 66; Mismatches 74; Indels 0; Gaps 0;
QY 1 MNLVNAQLOKQVYKFIQDEYVAILNLEEEYHNMSSESVVEKYLKLDIINLTNYL 60
Db 1 MSLVNRKQLEKXANVRFTQDEYVAILDALEEEYHNMSSENTVVEKYLKLDIINLTDIYI 60
QY 61 NTKYKSGRNKALKKFKKEYLVTEVLEKNNLTPVEKNLHFVWIGQINDTAINYNQWKD 120
Db 61 DTYKSGRNKALKKFKKEYLVTEVLEKNNLTPVEKNLHFVWIGQINDTAINYNQWKD 120
QY 121 VNSDYTVKVPYDSNAFLINTLTKTIVESATNTLTLESFRENLDNPFDPYKFKRMEIYY 180
Db 121 VNSDYVNVFYDSNAFLINTLTKTVVESAINDTLESFRENLDNPFDPYKFKRMEIYY 180

```

QY 181 DKQHFIDYKSOIEENPEFIIDNIITKYLISNEYSKDLALNKYIEESLNKITANNNDI 240
Db 181 DKQNFINYKAQREENPELIIDDIIVKTYLSNEYSKEIDELNTYIEESLNKITQNSNDV 240
QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVILPGIOPDLFKS 300
Db 241 RNFBFEKNGESFNLYEQLVERWNLAAASDILRISALKEIGGMVYLDVMDLPGIOPDLFES 300
QY 301 INKPDSTINTSWEMIKLEAIMKYKEYIPEYTSKFNFDMLDEEVQSFESALSSKSDKSEIF 360
Db 301 IEKPSSTVTDVFWEMTKLEAIMKYKEYIPEYTSKFNFDMLDEEVQSFESALSSKSDKSEIF 360
QY 361 LPLDDIKVSPLEVKIAFANNVINQALISLSDKSCSDLVINQIKRYKILNDNLNPSINE 420
Db 361 SSLGMEASPLEVKIAFANNKGIINOGLISVSDKSCNLIIVKQIENRYKILNNSLNPAISE 420
QY 421 GTDFNTWKIPSDKLAISNEDNMFMKITNYILKVGFPAPDVRSTINLSGPGVYTGAYQD 480
Db 421 DNDFTNTTNTFIDSIMAEANADNGRFMMELGKYLVRGFPDVKTTINLSGPEAYAAAYQD 480
QY 481 LLMFKDNSTNHLLEPELRNPEFPKTKISOLTEQELTSLWSFNQARAKSQFEEYKKGFFE 540
Db 481 LLMFKEGSNVHLLEADURNFEISKTNISQTEQEMASLWSFDDARAKAQFEEYKRNFFE 540
QY 541 GALGEDDNLDFQANTV 556
Db 541 GSLGEDDNLDFSQNV 556

RESULT 10
US-10-463-957-9
; Sequence 9, Application US/10463957
; Publication No. US20040028705A1
; GENERAL INFORMATION:
; APPLICANT: Ballard, Jimmy D.
; APPLICANT: Spyres, Lea M.
; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637
; CURRENT APPLICATION NUMBER: US/10/463,957
; CURRENT FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: 60/389,685
; PRIOR FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 9
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-10-463-957-9

```

```

Query Match 39.5%; Score 2051.5; DB 12; Length 522;
Best Local Similarity 73.8%; Pred. No. 2e-121;
Matches 385; Conservative 66; Mismatches 70; Indels 1; Gaps 1;

QY 35 HNMSSSVVEKYLKLDINNLDTNLYNTYKSGRNKALKKFKYELTMEVLELKNNSLTPV 94
Db 1 HNMSENTVVEKYLKLDINSLDITDYIDTYKSGRNKALKKFKYELVIELEENSLTPV 60
QY 95 EKNLHFIWIGQNDTAINYNQKDVNSDYTVKVFYDSNAFLINTLTKKTIVESATNNTL 154
Db 61 EKNLHFIWIGQNDTAINYNQKDVNSDYTVKVFYDSNAFLINTLTKKTIVESASNDTL 120
QY 155 ESFRENLDPEFDYKFKRMBIYDKQGHFIDYKSOIEENPEFIIDNIITKYLISNEY 214
Db 121 ESFRENLDPEFNHTAFPRKMQIYYDKQNFINYKAQREENPDLIIDDIVKTYLSNEY 180
QY 215 SKOLEALNKYIEESLNKITANNNDIRNLEKFADEDLVRLYNQELVERWNLAAASDILRI 274
Db 181 SKOIDEALNAYIEESLNKVTNSGNDVRNFEFTGEVFNLYEQLVERWNLACASDILRV 240
QY 275 SMLKEDGGVYLDVILPGIOPDLFKSINKPDSI-TNTSWEMIKLEAIMKYKEYIPEYTSK 333
Db 241 AILKNTGGVYLDVMDLPGIHPDLFKINKPDSVKTAVDWEEMQLEAIMKHKEYIPEYTSK 300

```

```

QY 334 NFDMLDEEVQSFESALSSKSDKSEIFLPLDDIKVSPLEVKIAFANNVINQALISLSDK 393
Db 301 HFDTLDEEVQSFESVLAASKSEIFLPLGDIIVSPLEVKIAFAKGIINQALISAKDS 360
QY 394 YCSDLVINQIKRYKILNDNLNPSINEGTDFNTWKIPSDKLAISNEDNMFMKITNY 453
Db 361 YCSDLLIKQIQNRYKILNDTLGPIISQNDNTNNTWNGESLGAIAENENTSFAGISY 420
QY 454 LKVGFPAPDVRSTINLSGPGVYTGAYQDILLMFKDNSTNHLLEPELRNPEFPKTKISQLTE 513
Db 421 LAVGFPPEANTVTLSGPTIYAGAYKDLITFKEMSIDTSLSELNPEFPKVNISQATE 480
QY 514 QBITSLWSFNQARAKSQFEEYKKGIFEGALGEDDNLDFQANT 555
Db 481 QEKNSLWQNEBRAKIQFEEYKKNYFEGALGEDDNLDFSQNT 522

RESULT 11
US-10-463-957-3
; Sequence 3, Application US/10463957
; Publication No. US20040028705A1
; GENERAL INFORMATION:
; APPLICANT: Ballard, Jimmy D.
; APPLICANT: Spyres, Lea M.
; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637
; CURRENT APPLICATION NUMBER: US/10/463,957
; CURRENT FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: 60/389,685
; PRIOR FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-10-463-957-3

```

```

Query Match 38.1%; Score 1979; DB 12; Length 500;
Best Local Similarity 75.0%; Pred. No. 7.2e-117;
Matches 375; Conservative 60; Mismatches 65; Indels 0; Gaps 0;

QY 1 MSLVNRKAQLOKQVYKFRIOEDEYVAILNALBEEYHNMSSSVVEKYLKLDINNLDTNLY 60
Db 1 MSLVNRKQLEKMANVFRIOEDEYVAILDALBEEYHNMSENTVVEKYLKLDINSLDIYI 60
QY 61 NTKYKSGRNKALKKFKYELTMEVLELKNNSLTPVEKNLHFIWIGQNDTAINYNQKWD 120
Db 61 DTKYKSGRNKALKKFKYELVTEVLELKNNSLTPVEKNLHFIWIGQNDTAINYNQKWD 120
QY 121 VNSDYTVKVFYDSNAFLINTLTKKTIVESATNNTLESFRENLDPEFDYKFKRMBIY 180
Db 121 VNSDYNVNFYDSNAFLINTLTKKTIVESAINDTLESFRENLDPRFDYKFKRMBIY 180
QY 181 DKQGHFIDYKSOIEENPEFIIDNIITKYLISNEYSKDLALNKYIEESLNKITANNNDI 240
Db 181 DKQNFINYKAQREENPELIIDDIIVKTYLSNEYSKEIDELNTYIEESLNKITQNSNDV 240
QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVILPGIOPDLFKS 300
Db 241 RNFEFKNGESFNLYEQLVERWNLAAASDILRISALKEIGGMVYLDVMDLPGIOPDLFES 300
QY 301 INKPDSTINTSWEMIKLEAIMKYKEYIPEYTSKFNFDMLDEEVQSFESALSSKSDKSEIF 360
Db 301 IEKPSSTVTDVFWEMTKLEAIMKYKEYIPEYTSKFNFDMLDEEVQSFESVLAASKSDKSEIF 360
QY 361 LPLDDIKVSPLEVKIAFANNVINQALISLSDKSCSDLVINQIKRYKILNDNLNPSINE 420
Db 361 SSLGMEASPLEVKIAFANNKGIINOGLISVSDKSCNLIIVKQIENRYKILNNSLNPAISE 420
QY 421 GTDFNTWKIPSDKLAISNEDNMFMKITNYILKVGFPAPDVRSTINLSGPGVYTGAYQD 480

```



```
Db 421 DNDFTNTTTFIDSIMAEANADNGRFMBELGKLYRVGFPPDVKTITNLSGPEAYAAAYCD 480
Qy 481 LLMFKDNSTNIHLLEPLRN 500
Db 481 LLMFKEGSMNIHLLEADLRN 500

RESULT 12
US-10-463-957-5
; Sequence 5, Application US/10463957
; Publication No. US20040028705A1
; GENERAL INFORMATION:
; APPLICANT: Spvires, Lea M.
; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637
; CURRENT APPLICATION NUMBER: US/10/463,957
; PRIOR FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: 60/389,685
; PRIOR FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 5
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-10-463-957-5

Query Match 33.4%; Score 1734; DB 12; Length 420;
Best Local Similarity 78.3%; Pred. No. 1.7e-101;
Matches 329; Conservative 48; Mismatches 43; Indels 0; Gaps 0;

Qy 1 MNLVNAQLOKQVYKFRIOEDYVAIINALAEYHNMSSESVVEKYKDKDINNTDNYL 60
Db 1 MSLVNRKQLEKMANVRFTQEDYVAILDALAEYHNMSSESVVEKYKDKDINSLTDIYI 60
Qy 61 NTYKSGRNKALKKFEYLTMVELEKNNSLTPVEKNLHFPIWIGGOINDTAINYNQWKD 120
Db 61 DTYKSGRNKALKKFEYLTMVELEKNNSLTPVEKNLHFVWIGGOINDTAINYNQWKD 120
Qy 121 VNSDYTVKVFYDSNAFLINTLTKTIVESATNTLSEFRENLDNPPEDYKFKMEIY 180
Db 121 VNSDYNVNFYDSNAFLINTLTKTIVESAINDTLESFRENLDNPPEDYKFKMEIY 180
Qy 181 DKQKHFDIYKSOIEENPEFIIDNIITKTVLSNEYSKDLALNKYIEESLNKITANGNDI 240
Db 181 DKQKNFNYKQAEENPELLIIDIVKTVLSNEYSKEIDELNTYIEESLNKITQNSGNDV 240
Qy 241 RNLEKFADEDLVLYNQELVERWNLAASDILRISMLKSDGGVYLDVILPGIQLPKFS 300
Db 241 RNFEFPKNGESFNLYEDELVERWNLAASDILRISALKSIGWYLDVDMPLGQIPLPES 300
Qy 301 INKPDSTNTSWEMIKLEAIMKYEIPGYTSKNFMDLDEEVQSFESALSSKSKSEIF 360
Db 301 IEPSSVTVDVEMTKLEAIMKYEIPGYTSKHFMDLDEEVQSFESVLSKSKSEIF 360
Qy 361 LPDDDIKVSPLYKIAFANNSVNOALISLSDYSCLVINOIKNRYKILNDNLNPSINE 420
Db 361 SSLGDWEASPLEVKIAFNKSGIINQGLISVSDSYCSNLIVKQIENRYKILNLSNLPASE 420

RESULT 13
US-10-463-957-7
; Sequence 7, Application US/10463957
; Publication No. US20040028705A1
; GENERAL INFORMATION:
; APPLICANT: Spvires, Lea M.
; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637
; CURRENT APPLICATION NUMBER: US/10/463,957
; CURRENT FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: 60/389,685
```

```
; PRIOR FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-10-463-957-7

Query Match 14.8%; Score 769; DB 12; Length 170;
Best Local Similarity 85.3%; Pred. No. 4.7e-41;
Matches 145; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MNLVNAQLOKQVYKFRIOEDYVAIINALAEYHNMSSESVVEKYKDKDINNTDNYL 60
Db 1 MSLVNRKQLEKMANVRFTQEDYVAILDALAEYHNMSSESVVEKYKDKDINSLTDIYI 60
Qy 61 NTYKSGRNKALKKFEYLTMVELEKNNSLTPVEKNLHFPIWIGGOINDTAINYNQWKD 120
Db 61 DTYKSGRNKALKKFEYLTMVELEKNNSLTPVEKNLHFVWIGGOINDTAINYNQWKD 120
Qy 121 VNSDYTVKVFYDSNAFLINTLTKTIVESATNTLSEFRENLDNPPEDYK 170
Db 121 VNSDYNVNFYDSNAFLINTLTKTIVESAINDTLESFRENLDNPPEDYK 170

RESULT 14
US-10-114-170-257
; Sequence 257, Application US/10114170
; Publication No. US20030023075A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/114,170
; FILING DATE: 01-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/453,702
; FILING DATE: 03-DEC-1999
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 257:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3169 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```

SEQUENCE DESCRIPTION: SEQ ID NO: 257;		US-10-114-170-257
Query Match	8.13; Score 422.5; DB 14; Length 3169;	
Best Local Similarity	21.04; Pred. No. 2.2e-17;	
Matches 258; Conservative 197; Mismatches 450; Indels 323; Gaps 54;		
QY	14 YVKFERI---QDEYVAILNAL--EEYHNMBSVVVKYKLKDINLNTDNYLTKKSGR	68
DB	160 YIKIRKRGAEQDTTQSLINELLNGVDNRNTI--PFQKISELNDIHSYENQIKNSR	218
QY	69 N--KALKKFEYLWVLELKNSS-----LTPVKNLHFIMWGGQ	106
DB	219 KGIELVKQGLSSLLNDKNGKQLSDNASKIINLLAGIEVQSHKVDIEPFIHVVVAGA	278
QY	107 INDRAINVQKVDNSDYTVKVFYDSNAP-----LINTLTKT-----	144
DB	279 PPDNTFSYITAFNTYKDYTYLLWIDPNAFGAARFSGILKNIAMNYAIMELRNPHLAE	338
QY	145 -----IVESATNNTLE--SFRENLDPEFY-----NKPYRKRMEIIVDKKHFI	187
DB	339 EMNEVLKIQNIQNETIEFKETRELRKLEENRYKSLTSEYKTEKENVFFLESIMGMQDNYF	398
QY	188 DYKSOIEENPEFI-----IDNIUK--TYLSNEYSKOLEALNKYIEESLNKITANNGN--	238
DB	399 TYCISNGISNTDDISRLDFLNLVKLSPEVQNDKSVKKNKRDIDLLKNTISQKGDGF	458
QY	239 ---DIRNLEKFADELDVLYNQELVERNLAAASDIIRISMLKEDGGVYLDVDILPGIQP	295
DB	459 QLDRINTLESFKPKQDYFFYQOEMLLRWYAAASQVRLINLKKEYGGIYTDITDILPAYSD	518
QY	296 DLFKSINKPDSITNTSWEMIKLEAIMKYK--EYIPG--YTSKNFMDLDEEVQSFESALS	351
DB	519 KVSQIINE--KSDKRFEDLRLRIISSEIISLTKGEKYSIKH--DGLDETTLNQLNML--	575
QY	352 SKSDKSEIFLPUDDIKVSPLEVKVAFANNSVI-----NQALISLKDSSYC	395
DB	576 SEIEK-----LTDDY--FKPVETKVVRDFKIFKRYOKWTENTWIRGNNEFMTHKSKC	630
QY	396 SDLVINQIKNRYKILNDNLNPSINEGDTNTMKIFSCLKASISNE-----DNM	444
DB	631 IDFILSGQKKQYLE--QRARDNISYNNLFYIT-----EDLKSANVAIGGIPAKKYLEHG	684
QY	445 MPWKIKITNLYKVGFPADVRSTNLSP-----	471
DB	685 LP-----SEYRQDGTIPYVSTNLISGPDIMFQMKYKSLGRIGEVHAIKONKLSVDNVL	740
QY	472 GYVTGAYQDLMF-----KONSTNIHLLEP-----LRNFEPPKTKISOLT-----BOEFTS	518
DB	741 GYVASSNKDNKSFNMLNPVSVGINDITPDDESSWVRNNDINKILFEKINCHVPEKLPTS	800
QY	519 LWSFNQARAKSOFEEYKKGFFEGALGEDDNLDFACQNTVLDKDYVSKKILSSNMKTRNKEYI	578
DB	801 LY-----YEIDSRSPFGQW-----DNKSIKHVTIINKDLI--KDINLLLTSSNIDV	844
QY	579 HTVLOQDKIYBAYASCNLFKDPYSSILYQKNIEGSETAYYYYVADAEIKEDKIRIPY	638
DB	845 KLLIKL--DRELYAISKKI-----DNPLALSIRTLQOLANYVTSNTPEPENTINFIY	896
QY	639 QJNKRN-----IKLFTGHGKSEFNTDTFANLDVS-----LSSELET	677
DB	897 DFYRKQDQLLSAHL-----FSRNDADTKIIVYNSVMKNVFLSEVISC	942
QY	678 ILNLAKADISPKYIENLLNCNMFYSISAETYPGKLLKIKDRVSELMPSISQDSI--	735
DB	943 VLRSKKVD-----SVINEN-----KKNLSKEA--GALRDYAKLQKELFSLMLDDGVYK	990
QY	736 TVSANQYEVRAINPEGKEILHDHSGKWINKEESIIKDISKEYISFNP-----K	783
DB	991 IITNAY---IKERDKL-----SGIYNIENSIISGHESFDIIRGNQHEWGLSTVEQFK	1042
QY	784 ENKIIVGSKYLHELSTLQSIIRNANSODIDLEKKVNLTECEINVAS-----NIDRQIVE	838

RESULT 15
US-10-282-122A-76865
; Sequence 76865, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangseu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 76865
; LENGTH: 4688
; TYPE: PRT
; ORGANISM: Ureaplasma urealyticum
US-10-282-122A-76865

Query Match 5.6%; Score 290.5; DB 12; Length 4688;
Best Local Similarity 18.8%; Pred. No. 8.5e-09;
Matches 252; Conservative 220; Mismatches 433; Indels 437; Gaps 65;

```

1408 -----NNQOSLISAKIEVDDIDNVLTADKNIVYQ-----LENSDANNLKLATN 1454
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy 878 L-----DDSHFISFEDI-----SKTENGFRIRFINKETGNSIF----- 910
      : : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1455 AQVVVNNKKFLKFLDLVSLKINQYVKEISFNSKPTNAY-FNFTNKTNNIVSYDEQN 1513
      : : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy 911 -----ITEKEIF-SEYATHISKEISNIK-----DT 935
      : : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1514 KISLSNNINPTSKPOTINKDKSVNIDVLDQVQOLLANQYLRLLKQLNDNKTVMWTD 1573
      : : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy 936 IFDNVNGKLVKYNLDAAEHVNTLSAFTQSLI-----EYNTTKESLSN-LSVAMKQVYA 991
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1574 ILENNNAKISFKUS-NLIH-----NRAYELEGLYPDDQNSVNDMTNNQISFNSKIHKPK 1627
      : : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Qy 992 QLFSTGLNTIT-DASKVVELVS 1012
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1628 IEPESLTINYDNTNAIKTVS 1649
      : : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

Search completed: April 1, 2004, 16:54:32
Job time : 38.125 secs

```

Search completed: April 1, 2004, 16:54:32
Job time : 38.125 secs

REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-604A-10

Query Match 79.0%; Score 4101; DB 1; Length 2366;
Best Local Similarity 77.6%; Pred. No. 8.4e-238;
Matches 792; Conservative 106; Mismatches 122; Indels 0; Gaps 0;

QY 1 MNLVKAQLOKQVYKRIQDEYVAILNALBEEYHNMSESVVEKYKLDKINLTDNYL 60
DB 1 MSLVNRKQLEKXANVRFTQDEYVAILDALBEEYHNMSENTVVEKYKLDKINSLTDIYI 60

QY 61 NTKYKSGRNLKAKKFKKEYLTMEVLELKNLSLTPVEKNLHFWIGQINDTAINYNQWKD 120
DB 61 DTYKSGRNLKAKKFKKEYLTVELELKNLSLTPVEKNLHFWIGQINDTAINYNQWKD 120

QY 121 VNSDYTVKVFYDSNAFLNTLTKTIVESAATNTLESFRENLDNDFPDYKPFYKRMELIY 180
DB 121 VNSDYNVVFYDSNAFLNTLTKTIVESAINDTLESFRENLDNDFPDYKPFYKRMELIY 180

QY 181 DKQHFIDYKSGIEENFEITDNIITKYLSEYKSDLEALNKYIEBSLNKITANNNGDI 240
DB 181 DKQGNFINKYKAQREENPELIDIDIVKTYLSNEYSKSIDELNTVIESLNKITQNSGNDV 240

QY 241 RNLEKFADEDLVRNLQELVERWNLAAASDILRISMLKEDGGVLDVDIILPGIOPDLFKS 300
DB 241 RNFEFEKNGESFNLYEQELVERWNLAAASDILRISALKETGGMYLDVDMPLPGIOPDLFES 300

QY 301 INKPSITNTSWEMIKLEAMKYKEYIPGYTSKNFDMLDERVQSPFESALSSKSKSEIF 360
DB 301 IEKSSVTVDPEWMTKLEAMKYKEYIPEYTSERFMDLDEEVQSPFESVLASKSKSEIF 360

QY 361 LPLDDIKVSPLEVKIAPANNVINQALISLKDSCSDIVINQIKNRYKILNDNLNPSINE 420
DB 361 SSLGDMELASPLEVKIAFNKGIHQGLISVKDSCSNLIVKQIENRYKILNNSLNPAISE 420

QY 421 GTDFTNTWKIPSDKLASISNEDNMFMKITYNLKVGFPADVRSTINLSGGVYTGAYQD 480
DB 421 DNDFTNTTTFIDSIMAEANADNGRFMMELGKLYRVGFPDPVKTTINLSGPEAVAAAYQD 480

QY 481 LLMFKDNTNHLLEPELANFEFPKTKISQTEQBITSLASFPNOARAKSQPEEYKKGYPF 540
DB 481 LLMFKEGSMNHLIEADLRNFEISKTNTSQTEQEMASLWSFDDARAKAQPEEYKKNYFE 540

QY 541 GALGEDDNLDPANVTLDKDYVSKKILSSMKTRNKEYIHYIYVQLQGDKISYEASCNLFSK 600
DB 541 SSLGEDDNLDPQNVLDKDYVLEKISSLASRSGYIHYIYVQLQGDKISYEACNLPFK 600

QY 601 DPYSILYQKNIEGSETAYYYVADAEIKEDKRYIPYQISNKNIKLTFIGHGKSFBNT 660
DB 601 TPYDSVLQKNLEDESEIAYVYNGPGEIQTEDKYPISIIDRPKIKLTFIGHGKDFEFT 660

QY 661 DTFANLDVDSGSSETITLINAKADISPKYTEINLLGCNMFYSISABETYPGKLLKIK 720
DB 661 DIFAGFDVDSLSTEIAAIDLAKEDISPKSTEINLLGCNMFYSINVEETYPGKLLKVK 720

QY 721 DAVSILMPSISODSITVSANQYEVINEBGRKREILDHSGKWNKEESIKDISKEYISF 780
DB 721 DKRISLMPISODSITVSANQYEVINEBGRREILDHSGWINKESIKDISKEYISF 780

QY 781 NPKENKIIVKSKYLHELSTLLOEIRNANSDDILEKVMLTECHINVASNIDQIYVEGR 840
DB 781 NPKENKIIVKSKNLPSTLLOEIRNNSDDILEEKVMLTECHINVISNIDQIYVEER 840

QY 841 IBEAKNLTSDSINYIKNEFKLIESDSLVDLKHQNGLDSDHSHFISFEDISKTENGFRIRF 900

RESULT 2

US-08-405-496A-10
Sequence 10, Application US/08405496A
Patent No. 5919665

GENERAL INFORMATION:

APPLICANT: WILLIAMS, JAMES A.
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
TITLE OF INVENTION: NEUROTOXIN
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/405.496A
FILING DATE: 16-MAR-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/329,154

FILING DATE: 25-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/161,907

FILING DATE: 02-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/985,321

FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/429,791

FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: INGOLIA, DIANE E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPHD-01308

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 2366 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-405-496A-10

Query Match 79.0%; Score 4101; DB 2; Length 2366;

Best Local Similarity 77.6%; Pred. No. 8.4e-238;

Matches 792; Conservative 106; Mismatches 122; Indels 0; Gaps 0;

QY 1 MNLVKAQLOKQVYKRIQDEYVAILNALBEEYHNMSESVVEKYKLDKINLTDNYL 60

DB 1 MSLVNRKQLEKXANVRFTQDEYVAILDALBEEYHNMSENTVVEKYKLDKINSLTDIYI 60

QY 61 NTKYKSGRNLKAKKFKKEYLTMEVLELKNLSLTPVEKNLHFWIGQINDTAINYNQWKD 120

Db 61 DTYKSGRNKALKKFEYLTVLELKNLTPVEKNLHFVWIGGOINDTAINYNQWKD 120
Qy 121 VNSDYTVKPYDSNAFLINTLKTIVESAATNTLESFRENLDPEFDYKFKRMEIYY 180
Db 121 VNSDYNNVYDSNAFLINTLKTIVESAINDTLESFRENLDPRFDYKFKRMEIYY 180
Qy 181 DKQGHFDYKSGQIEENPEFIIONIKTYLSNYSKDEALNKYIEESLNKITTANNNDI 240
Db 181 DKQKNFINYKQARENPELIIDIVKTYLSNYSKEIDELNTYIEESLNKITQNSGNDV 240
Qy 241 RNLEKFADEDLVRLYNOLVERNLAAASDILRAISMLKEDGCVYLDVILPGIOPDLFXS 300
Db 241 RNLEKFADEDLVRLYNOLVERNLAAASDILRAISMLKEDGCVYLDVILPGIOPDLFXS 300
Qy 301 INKPDSTINTSWEMIKLEALIMKYKEYIPGYTSKNFMDLDEVOQSPESALSSKSDSEIF 360
Db 301 IEPSSVTVDFWEMTKLEALIMKYKEYIPEYTSSEHFDMLDEVOQSPESVLSKSDSEIF 360
Qy 361 LPDDIKVPLEVVKIAPANNVINQALISKDSYCSOLVINQIKNRYKILNDNLNPSINE 420
Db 361 SSLGDMESPLEVKIAPANNVINQALISKDSYCSOLVINQIKNRYKILNDNLNPSINE 420
Qy 421 GTDFNTMTKIFSDKLASINEDNMFMKLTNYLVKGFADPVDSTINLSPGYTTCAYOD 480
Db 421 DNDFTNTTTFIDISMAEADNADGRFMWELGKYLVRGFFDPVKTINLSGPEYAAAYQD 480
Qy 481 LLMEKDNSTNHLPELRLNPEFPKTKISQUTQEITSLMSFNQARAKSOFEEYKKGYPE 540
Db 481 LLMEKDNSTNHLPELRLNPEFPKTKISQUTQEITSLMSFNQARAKSOFEEYKKGYPE 540
Qy 541 GALGEDNLPQAQTVLDKDVSKILSSMKTRNKEYIHYIVOLQDGKISYEASCNLFSK 600
Db 541 GSLGEDNLPQAQTVLDKDVSKILSSMKTRNKEYIHYIVOLQDGKISYEASCNLFSK 600
Qy 601 DPYSILYOKNIESEFAYVYVADAEIKIDKIRIPYQISNKNKILFIHCKSEFNT 660
Db 601 TPYDSVLFOKNIESEFAYVYVADAEIKIDKIRIPYQISNKNKILFIHCKSEFNT 660
Qy 661 DTFANLVDLSSEIETILNAKADISPKYIEINLGCNMFYSISAEETYPCKLLKIK 720
Db 661 DTFANLVDLSSEIETILNAKADISPKYIEINLGCNMFYSISAEETYPCKLLKIK 720
Qy 721 DRVSELMPSISQDSITVSANQYEVRIEENEGKREILDSHGKWKINKEESIIKDISKEYISF 780
Db 721 DRVSELMPSISQDSITVSANQYEVRIEENEGKREILDSHGKWKINKEESIIKDISKEYISF 780
Qy 781 NPKENKIIVSKYLHELSTLLOETIRNANSDDIDLEKKVMTCEINVASNIDROQIVEGR 840
Db 781 NPKENKIIVSKYLHELSTLLOETIRNANSDDIDLEKKVMTCEINVASNIDROQIVEGR 840
Qy 841 IEEAKNLTSDSINIKNEFKLIESISDSLYDLKHONGLDSDHSFISPEDISKTEGPRIF 900
Db 841 IEEAKNLTSDSINIKNEFKLIESISDSLYDLKHONGLDSDHSFISPEDISKTEGPRIF 900
Qy 901 INKGTGNSIFTEKEKTEFSEVATHISKEISNIDKTIPTDNVNGKLVKKNLDAAEVNTLN 960
Db 901 INKGTGNSIFTEKEKTEFSEVATHISKEISNIDKTIPTDNVNGKLVKKNLDAAEVNTLN 960
Qy 961 SAFTIQLSILEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
Db 961 SAFTIQLSILEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020

RESULT 3
US-08-915-136-10
; Sequence 10, Application US/08915136
; Patent No. 6290360
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.

APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESS: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,136
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/480,604
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-915-136-10

Query Match 79.0%; Score 4101; DB 3; Length 2366;
Best Local Similarity 77.6%; Pred. No. 8.4e-238;
Matches 792; Conservative 106; Mismatches 122; Indels 0; Gaps 0;
Qy 1 MSLVNRKQLEKXVYKFRQDEYVAILNDALEEVHNMSENTPVEKNLHFVWIGGOINDTAINYNQWKD 60
Db 1 MSLVNRKQLEKXVYKFRQDEYVAILNDALEEVHNMSENTPVEKNLHFVWIGGOINDTAINYNQWKD 60
Qy 61 NTKYKSGRNKALKKFEYLTVLELKNLTPVEKNLHFVWIGGOINDTAINYNQWKD 120
Db 61 NTKYKSGRNKALKKFEYLTVLELKNLTPVEKNLHFVWIGGOINDTAINYNQWKD 120
Qy 121 VNSDYTVKPYDSNAFLINTLKTIVESAATNTLESFRENLDPEFDYKFKRMEIYY 180
Db 121 VNSDYTVKPYDSNAFLINTLKTIVESAATNTLESFRENLDPEFDYKFKRMEIYY 180
Qy 181 DKQGHFDYKSGQIEENPEFIIONIKTYLSNYSKDEALNKYIEESLNKITTANNNDI 240
Db 181 DKQGHFDYKSGQIEENPEFIIONIKTYLSNYSKDEALNKYIEESLNKITTANNNDI 240
Qy 241 RNLEKFADEDLVRLYNOLVERNLAAASDILRAISMLKEDGCVYLDVILPGIOPDLFXS 300
Db 241 RNLEKFADEDLVRLYNOLVERNLAAASDILRAISMLKEDGCVYLDVILPGIOPDLFXS 300
Qy 301 INKPDSTINTSWEMIKLEALIMKYKEYIPGYTSKNFMDLDEVOQSPESALSSKSDSEIF 360
Db 301 IEPSSVTVDFWEMTKLEALIMKYKEYIPEYTSSEHFDMLDEVOQSPESVLSKSDSEIF 360
Qy 361 LPDDIKVPLEVVKIAPANNVINQALISKDSYCSOLVINQIKNRYKILNDNLNPSINE 420
Db 361 LPDDIKVPLEVVKIAPANNVINQALISKDSYCSOLVINQIKNRYKILNDNLNPSINE 420
Qy 421 GTDFNTMTKIFSDKLASINEDNMFMKLTNYLVKGFADPVDSTINLSPGYTTCAYOD 480
Db 421 GTDFNTMTKIFSDKLASINEDNMFMKLTNYLVKGFADPVDSTINLSPGYTTCAYOD 480
Qy 481 LLMEKDNSTNHLPELRLNPEFPKTKISQUTQEITSLMSFNQARAKSOFEEYKKGYPE 540
Db 481 LLMEKDNSTNHLPELRLNPEFPKTKISQUTQEITSLMSFNQARAKSOFEEYKKGYPE 540
Qy 541 GALGEDNLPQAQTVLDKDVSKILSSMKTRNKEYIHYIVOLQDGKISYEASCNLFSK 600
Db 541 GALGEDNLPQAQTVLDKDVSKILSSMKTRNKEYIHYIVOLQDGKISYEASCNLFSK 600
Qy 601 DPYSILYOKNIESEFAYVYVADAEIKIDKIRIPYQISNKNKILFIHCKSEFNT 660
Db 601 DPYSILYOKNIESEFAYVYVADAEIKIDKIRIPYQISNKNKILFIHCKSEFNT 660
Qy 661 DTFANLVDLSSEIETILNAKADISPKYIEINLGCNMFYSISAEETYPCKLLKIK 720
Db 661 DTFANLVDLSSEIETILNAKADISPKYIEINLGCNMFYSISAEETYPCKLLKIK 720
Qy 721 DRVSELMPSISQDSITVSANQYEVRIEENEGKREILDSHGKWKINKEESIIKDISKEYISF 780
Db 721 DRVSELMPSISQDSITVSANQYEVRIEENEGKREILDSHGKWKINKEESIIKDISKEYISF 780
Qy 781 NPKENKIIVSKYLHELSTLLOETIRNANSDDIDLEKKVMTCEINVASNIDROQIVEGR 840
Db 781 NPKENKIIVSKYLHELSTLLOETIRNANSDDIDLEKKVMTCEINVASNIDROQIVEGR 840
Qy 841 IEEAKNLTSDSINIKNEFKLIESISDSLYDLKHONGLDSDHSFISPEDISKTEGPRIF 900
Db 841 IEEAKNLTSDSINIKNEFKLIESISDSLYDLKHONGLDSDHSFISPEDISKTEGPRIF 900
Qy 901 INKGTGNSIFTEKEKTEFSEVATHISKEISNIDKTIPTDNVNGKLVKKNLDAAEVNTLN 960
Db 901 INKGTGNSIFTEKEKTEFSEVATHISKEISNIDKTIPTDNVNGKLVKKNLDAAEVNTLN 960
Qy 961 SAFTIQLSILEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
Db 961 SAFTIQLSILEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020

QY 841 IBEAKNLTSDSINYIKNEFKLIESISDSLYDLKHONGLDSDSHFISPDISTKTENGPRIRF 900
DB 841 IEEAKNLTSDSINYIKNEFKLIESISDALCDLKQONELEDSDSHFISPDISTKTENGPRIRF 900
QY 901 INKETSIFETKEKEIFSEYATHISKEISNIDTIFDNVNGKLVKKVNLDAHEVNTLN 960
DB 901 INKETSIFETKEKEIFSEYATHISKEISNIDTIFDNVNGKLVKKVNLDTTHEVNTLN 960
QY 961 SAFFIQSLIEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
DB 961 AAFPIQSLIEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020

RESULT 6

US-09-084-517-10
Sequence 10, Application US/09084517
Patent No. 6613329
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: WILLIAMS, JAMES A.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESS: HAYSTACK, MEDLEN & CARROLL
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/084.517
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329.154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161.907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985.321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429.791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: OPND-01610
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-084-517-10

Query Match 79.0%; Score 4101; DB 4; Length 2366;
Best Local Similarity 77.6%; Pred. No. 8, 4e-238;
Matches 792; Conservative 106; Mismatches 122; Indels 0; Gaps 0;

QY 1 MNLVNAQOLQKVVYKFRIOEDYVAILNALVEEYHNSSESVVEKYLLKLDINNLTDNYL 60
DB 1 MSLVNRKQLEKMANVRFETQEDYVAILDALVEEYHNSSENTEVVEKYLLKLDINSLTDIYI 60
QY 61 NTKKSGKGNKALKKFKPEYLTVEVLELKNNSLTPVEKNLHFWIGQINDTAININQWKD 120
DB 61 DTKKSGKGNKALKKFKPEYLTVEVLELKNNSLTPVEKNLHFWIGQINDTAININQWKD 120
QY 121 VNSDYTVKVFYDSNAFLINTLTKTIVTESATNTTLESFRENLDPEFYKNPKYRMEIY 180
DB 121 VNSDYTVKVFYDSNAFLINTLTKTIVTESATNTTLESFRENLDPEFYKNPKYRMEIY 180
QY 181 DKQKHFIDYKSOIBENPEFIIDNIITKYLSEVSKOLEALNKYTEBSLNKITANGNDI 240
DB 181 DKQKHFIDYKSOIBENPEFIIDNIITKYLSEVSKOLEALNKYTEBSLNKITANGNDI 240
QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDLIRLSMLKEDGGVYLDVLDILGFIQDPLPKS 300
DB 241 RNLEKFADEDLVRLYNQELVERWNLAAASDLIRLSMLKEDGGVYLDVLDILGFIQDPLPKS 300
QY 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKQFMDLDEEVQSFESALSCKSDKSEIF 360
DB 301 INKPSITNTSWEMIKLEAIMKYKEYIPGYTSKQFMDLDEEVQSFESALSCKSDKSEIF 360
QY 361 LPLDDIKVSPLEVKIAFANNSVINQALISLKDSCDLVINQIKRKYKILNDNLNPSINE 420
DB 361 LPLDDIKVSPLEVKIAFANNSVINQALISLKDSCDLVINQIKRKYKILNDNLNPSINE 420
QY 421 GTDFNTWKIPSDKLASISNEDNMFMKITYNLKVGFPADPVRSTINLSGGVYTCAYQD 480
DB 421 GTDFNTWKIPSDKLASISNEDNMFMKITYNLKVGFPADPVRSTINLSGGVYTCAYQD 480
QY 481 LMFKEGSNWNIHLIADLRNFEISKTINSQTEQWASLMSFDDARAKAQFEYKRNIFE 540
DB 481 LMFKEGSNWNIHLIADLRNFEISKTINSQTEQWASLMSFDDARAKAQFEYKRNIFE 540
QY 541 GALGEDDNLDFQNTVLDKDYVSKKILSSMKTRKEYIHYIYVQLQGDKISYEASCNLFSK 600
DB 541 GALGEDDNLDFQNTVLDKDYVSKKILSSMKTRKEYIHYIYVQLQGDKISYEASCNLFSK 600
QY 601 DPYSILYQKNTGSETAYVYVADAETKEIDKYRIPYQISNKNRNIKUTFTIGHGSEPT 660
DB 601 DPYSILYQKNTGSETAYVYVADAETKEIDKYRIPYQISNKNRNIKUTFTIGHGSEPT 660
QY 661 DTFANLDVDSLSSEIETILNLAADISPKYIBINLLGCMFSPYSISABETPGKLLLIK 720
DB 661 DTFANLDVDSLSSEIETILNLAADISPKYIBINLLGCMFSPYSISABETPGKLLLIK 720
QY 721 DRVSELMPISISODSITVSANOYEVEINEEGREILDHSGKWINKKEEIIKOISKEYISF 780
DB 721 DRVSELMPISISODSITVSANOYEVEINEEGREILDHSGKWINKKEEIIKOISKEYISF 780
QY 781 NPENKIIYKSKYLHELSTLLQEIERNANSSDIDLEKKVMTLCEINVASNIDRQIVEGR 840
DB 781 NPENKIIYKSKYLHELSTLLQEIERNANSSDIDLEKKVMTLCEINVASNIDRQIVEGR 840
QY 841 IBEAKNLTSDSINYIKNEFKLIESISDSLYDLKHONGLDSDSHFISPDISTKTENGPRIRF 900
DB 841 IEEAKNLTSDSINYIKNEFKLIESISDALCDLKQONELEDSDSHFISPDISTKTENGPRIRF 900
QY 901 INKETSIFETKEKEIFSEYATHISKEISNIDTIFDNVNGKLVKKVNLDAHEVNTLN 960
DB 901 INKETSIFETKEKEIFSEYATHISKEISNIDTIFDNVNGKLVKKVNLDTTHEVNTLN 960
QY 961 SAFFIQSLIEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
DB 961 AAFPIQSLIEYNTTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020

RESULT 7
US-08-480-604A-6
Sequence 6, Application US/08480604A

Patent No. 5736139

GENERAL INFORMATION:

APPLICANT: KINK, JOHN A.

APPLICANT: THALLEY, BRUCE S.

APPLICANT: PADHYE, NISHA V.

APPLICANT: FIRCA, JOSEPH R.

APPLICANT: STAFFORD, DOUGLAS C.

TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND PREVENTION OF C. DIFFICILE DISEASE

TITLE OF INVENTION: 32

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: UNITED STATES OF AMERICA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/480,604A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/422,711

FILING DATE: 14-APR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/405,496

FILING DATE: 16-MAR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/329,154

FILING DATE: 25-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/161,907

FILING DATE: 02-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/985,321

FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/429,791

FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: INGOLIA, DIANE E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPHD-01763

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 2710 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-480-604A-6

Query Match 53.8%; Score 2791; DB 1; Length 2710;
Best Local Similarity 52.1%; Pred. No. 4.6e-159;
Matches 534; Conservative 221; Mismatches 262; Indels 8; Gaps 5;

	QY	1	MNLVNKAOLQKQVYVKPRIQEDEVAILNLEEYHNHNSSSSVVEKYKLKLOINNUTDNYL	60
			: : : : : : : : : : : : : : : : : : : : : :	
	Db	1	MSLI8KEELIKLAY-SIRPRENEYKTILTLDENKLTITNNENKYQLQCLNESIDVFEM	59
			: : : : : : : : : : : : : : : : : : : : : :	
	QY	61	NTYKSGRNKALKKFKFYLTMVELLEKNSLTPVEKNLHFMTIGGQINDATAYINOKMD	120
			: : : : : : : : : : : : : : : : : : : : ~:	
	Db	60	NKYTSSNRNALSNLKDIKLEVLLIKNSTSPVEKNLHFMTIGGVSDIALEYTKQWAD	119
			: : : : : : : : : : : : : : : : : : : : ~:	
	QY	121	VNSDYTVKVFYDSNAFLINTLKTIIVSATNTTLGSRFNLDPEFDYNPKYRKMEIYY	180
			: : : : : : : : : : : : : : : : : : : : ~:	

Db	120	INAEYNIKUWDSBFLVNTLKAIVESSTTALQLEBEIQNPQDFNNKFKYKRMFEFY	179
Qy	181	DKQKHFDIYKQSBEPFIIDNIIKTYLSNEYKDLALNKYIEESLNKTIANNNDI	240
Db	180	DRQKRFINYKQINQKPTPTIDDIKSHLVSEYNRDETIVLESYRTNSLRKINSHGIDI	239
Qy	241	RULEKPADBDLVRVLNQELVERWNLAASDIIRISMLKEDGGVYLDVDLPGIQDPLFKS	300
Db	240	RANSLEFTEOELLNIYQCELLNRGNLAASDVRLLALKNFQGVYLDVMDLPGIHSDLFT	299
Qy	301	INKPDSITWTSWEMIKLEAIMKYKEYIPGVTSKNFDMLEEVQRPESALSXSDDKSEIF	360
Db	300	ISPSGIGDRWEMIKLEAIMKYKYNNTYSENFDKQDQKDNFKLIESEKSEIF	359
Qy	361	LPDIDKVSPLVKIAFANNKSVINQALISKDSYCSDLVINQIKRYKILNDNPESINE	420
Db	360	SKULENLVSDLEIKIAFALGSVINQALISKQSYLTNLVIEQVKNRYQFLNQHLPABIS	419
Qy	421	GTFDNTMKIFSDKLASINENDNMFMIKITNYLVGFAPDVRSTINLSPGPGVYTCAYOD	480
Db	420	DNNFTDTTKIFHDSLNSATAENSMPLTIAPLYQVGFMPPEARSTISLSGPGAYASAYD	479
Qy	481	LMFMKONSINIHLEPELNFNFPKTKISOLTEQBITSLWSFNQARAKSOFEEYKKGVEE	540
Db	480	FINLOQNTIEKTLKASDLIEFKFPPENNLSQLTEQIEINLSWSDQASAKYQFEKYVRDVTG	539
Qy	541	GALGEDNDLFAONTVLDXDY-VSKKILSS--MKTRNKEYIHVIVLOQDQKISYEASCNL	597
Db	540	GSLSGSDGVDFNKATDLDKXVLLNKKIPSNVVEAGSKNYVHYIQLQDQDLSYEATCNL	599
Qy	598	FSKDPYSSILYKQNIIEGSTAYYYVAD--ASIEKIDKYRIPIYQISNKNIKLTFIGHGK	655
Db	600	FSKNPKNSIIIQRM--NESAKSYFLSDDGESILELNKYRIPELKNKEKVKVTFIGHGK	657
Qy	656	SEBNTDTFANLDVDSLSEIETILNLAKADISPKYIEINLLGNNMPSYSISAEETYPGKL	715
Db	658	DEPNTSEFARLSVDSLUSNEISSFLDTIKUDISPKNVENVNLLGNNMPSYDFNVEETYPGKL	717
Qy	716	LLKIKDRVSELMPISQDSITVTSANQYEVINEEGKREILDHSGKWINKEESIIKDISK	775
Db	718	LLSMDKISTPDPVKNKSIITGANKQYEVINSEGEKELLASHGKWINKEEAIMSDLSK	777
Qy	776	EYISFNPKNENKIIVKSKYLHEISTLLOETIRNANSDDIDLEKKVMLTECEINVASNIDRQ	835
Db	778	EYIFPDSIDNKLKAKSKNIPGLASISEDIKTLLDASVSPDKFILNNLKLNISSIGDY	837
Qy	836	IYVGRITEEAKNLTSDSINYVKNFPLIESISDSLYDLKHONGLDSDHFTSFEDISKTENG	895
Db	838	IYYEKLEPVKNIIHNSIDBLIDEFNLLENVDELYELKKNLNDLDEKYLISFEDISKQNET	897
Qy	896	FRIRFNKGTGNSIFETEKEIFSEYATHISKEISNIKDTIFDNVNGKLVKKVNLDAAEH	955
Db	898	YSVRFINKNGBSVYVETEKEIFSKYSEHITKEISTKNSIITDVNGNLLDNIQLDHTSQ	957
Qy	956	VNTLNSAFFIQLIEYNTTKESLSLSVAMKYQVYAQLFSTGLNTITTDASKVVELVSTAL	1015
Db	958	VNTLNAAFFIQLISLDYSSNKNVDLNDLSTSVKQVYAQLFSTGLNTIYDSTQLVNLISNAV	1017
Qy	1016	DETID 1020	
Db	1018	NDTIN 1022	
RESULT 8			
US-08-405-496A-6			
; Sequence 6, Application US/08405496A			
; Patent No. 5919665			
; GENERAL INFORMATION:			
; APPLICANT: WILLIAMS, JAMES A.			
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM			
; TITLE OF INVENTION: NEUROTOXIN			
; NUMBER OF SEQUENCES: 30			
; CORRESPONDENCE ADDRESS:			

```

; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 16-MAR-1995
; APPLICATION NUMBER: US/08/405,496A
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-405-496A-6
;
; Query Match 53.8%; Score 2791; DB 2; Length 2710;
; Best Local Similarity 52.1%; Pred. No. 4.6e-159;
; Matches 534; Conservative 221; Mismatches 262; Indels 8; Gaps 5;
;
QY 1 MNLVKAQIQKVVYKFIQDEYVAIINLAEVEYHNMSESSVVEKYKLKIDNNLTNYL 60
DB 1 MSLISKELIKLAY-SIRPRENEYKITLTDENKLTNNENKYLQKLNESIDVFM 59
QY 61 NPYKSGRNKALKKPKFYLTWEVLKKNLSLTPVEKNLHFIWIGQINDTAINYQWKD 120
DB 60 NKYKTSRRNALSNLKKDILKEVLKNSNTSPVEKNLHFVWIGVEVSDIALEVIKQWAD 119
QY 121 VNSDYTVKVFYDSNAFLNTLTKTIVSATNTLESFRENLDPEFDYKPKRGMIYY 180
DB 120 INAEYIKLWYDSEAFNLTKKAIVESTTEALQLLEEIQNFQDNMKPKYKRMFEIY 179
QY 181 DKQHFIDYKSOIENPEFIDNIKTYLSNEYSKDLAENKYIEBSLKITANNNGDI 240
DB 180 DKQKFPYKQINKPVTPTIDDIKHLVSEYRDETVLSEYRTNLSLRKINSHGDI 239
QY 241 RLKLEFADBLVRLNQLVERWNLAASDILRISMLXEDGGVLDVLDILFGIQDLPFKS 300
DB 240 RANSLFTEQELNLYSQELLNRGNLAASDILVLLALNFGGVLDVLDMLFGIHSDLFT 299
QY 301 INKPSITNTSWEMIKLRAIMKYKFIYGYTSKNFMDLDEEVQSFBSALSKSKDKGRIF 360
DB 300 ISRPSIGLDRWEMIKLRAIMKYKFIYNNYTSNFDFKLDQQLKONFKLIIESKSEKGRIF 359
QY 361 LPDLIDKPSPLEVKAFANNSVINGALISLKDSCYSDLVINQIKRYKILANDNLPSPINE 420
DB 360 SKLENLNVSDLEIKIAFALGVSINGALISKQSYLTNLIVIEQVKNRYQFLNOLNPAIES 419
;
; ADDRESS: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 16-MAR-1995
; APPLICATION NUMBER: US/08/405,496A
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-405-496A-6
;
; Query Match 53.8%; Score 2791; DB 2; Length 2710;
; Best Local Similarity 52.1%; Pred. No. 4.6e-159;
; Matches 534; Conservative 221; Mismatches 262; Indels 8; Gaps 5;
;
QY 421 GTDFNTTKMFIKSFQKLASINEDNMFMKITNYLVKGFAPDVRSTINISGPGVYTGAYQD 480
DB 420 DNNFTDTTKI FHDLSFN SATAENSFLTKIAPYLQVGFMPPEARSTISLSPGAYASAYD 479
QY 481 LLMFKDNSTNIHLLEPELRNPFPPKTSQITQEQITSLWSPNQARAKSQBEYKKGTFE 540
DB 480 FINQENTTEKTAKASDLIEFKFPENNLSQLTEQEQINSLSWSPDQASAKYQPEKYVDYTG 539
QY 541 GAGEDDNLDFRQNTVLDKDY-VSKKILSS--MKTRNKYEHYIYQLOGDKISYEASCNL 597
DB 540 GSLSEDNGVDFNKNTALDKNYLLNNKIPSNVVEAGSKNYHYIIQLOGDDLSYATCNL 599
QY 598 FSKDPYSSILYQKNIETGETAYYYVAD--ABIKEIDKYRIPYQISNKRNIKLTFIGHGK 655
DB 600 FSKNPKNSIIQBNM--NESAKSYFLSDGSEILNKYRIPERLKNKVKVTFIGHGK 657
QY 656 SEFNTDTFANLVDVSLSSIEITILNAKADISPKYIEINLGCNMFPSYSISABEYTPGKL 715
DB 658 DEFNTSEFARLSVDSLSEISSEISFLDTIKLIDISPKVNEVNLGCNMFSDYFNVEETYPGKL 717
QY 716 LLKIKDRVSELMPISQDSIITVSANQYVRINEEGKREILDHSGKWINKKEESIIKDISK 775
DB 718 LLSIMDKITSTLFDVNNKNSITIGANQYVRINSEGRKELLASHSGKWINKKEEAIMSDLSK 777
QY 776 EYISFNPKENKIIKSKYLHELSTLLQELRNANSDDIDLEKKVMLTECEINVASNIDRQ 835
DB 778 EYIFFDSIDNKLKAKSKNIPGLASISEDIKTLILDASVSPDTKFLNNLKLNIIESIGDY 837
QY 836 IVEGRIIEAKNLTSDSINYKNEFKLIESISDLYDLKHQGLDSDSHFISFEDISKTENG 895
DB 838 IYVEKLEPVKNIHNSIDDLIDFNLLENVSDLYELKLNLDKYLISPEDISKNST 897
QY 896 FRIRPINKETGNSIFETKEIFSEYATHISKEINIKDITFDVNGKLVKKNLDAAHE 955
DB 898 YSVRFINKSGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGLLDNIQDHTSQ 957
QY 956 VNTLASAFIQSLIEYNTTKESLSNLSVAMKYQVLAQFLSTGLNTITDASKVVELVSTAL 1015
DB 958 VNTLNAAFIQSLIDYSSNKVDNLDSLTSVKVQLAQLFSTGLNTIYDSIQLVNLISNAV 1017
QY 1016 DETID 1020
DB 1018 NDTIN 1022
;
; US-08-915-136-6
; Sequence 6, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTIPOXIN FOR TREATMENT AND PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 16-MAR-1995
; APPLICATION NUMBER: US/08/405,496A
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-405-496A-6
;
; Query Match 53.8%; Score 2791; DB 2; Length 2710;
; Best Local Similarity 52.1%; Pred. No. 4.6e-159;
; Matches 534; Conservative 221; Mismatches 262; Indels 8; Gaps 5;
;
QY 1 MNLVKAQIQKVVYKFIQDEYVAIINLAEVEYHNMSESSVVEKYKLKIDNNLTNYL 60
DB 1 MSLISKELIKLAY-SIRPRENEYKITLTDENKLTNNENKYLQKLNESIDVFM 59
QY 61 NPYKSGRNKALKKPKFYLTWEVLKKNLSLTPVEKNLHFIWIGQINDTAINYQWKD 120
DB 60 NKYKTSRRNALSNLKKDILKEVLKNSNTSPVEKNLHFVWIGVEVSDIALEVIKQWAD 119
QY 121 VNSDYTVKVFYDSNAFLNTLTKTIVSATNTLESFRENLDPEFDYKPKRGMIYY 180
DB 120 INAEYIKLWYDSEAFNLTKKAIVESTTEALQLLEEIQNFQDNMKPKYKRMFEIY 179
QY 181 DKQHFIDYKSOIENPEFIDNIKTYLSNEYSKDLAENKYIEBSLKITANNNGDI 240
DB 180 DKQKFPYKQINKPVTPTIDDIKHLVSEYRDETVLSEYRTNLSLRKINSHGDI 239
QY 241 RLKLEFADBLVRLNQLVERWNLAASDILRISMLXEDGGVLDVLDILFGIQDLPFKS 300
DB 240 RANSLFTEQELNLYSQELLNRGNLAASDILVLLALNFGGVLDVLDMLFGIHSDLFT 299
QY 301 INKPSITNTSWEMIKLRAIMKYKFIYGYTSKNFMDLDEEVQSFBSALSKSKDKGRIF 360
DB 300 ISRPSIGLDRWEMIKLRAIMKYKFIYNNYTSNFDFKLDQQLKONFKLIIESKSEKGRIF 359
QY 361 LPDLIDKPSPLEVKAFANNSVINGALISLKDSCYSDLVINQIKRYKILANDNLPSPINE 420
DB 360 SKLENLNVSDLEIKIAFALGVSINGALISKQSYLTNLIVIEQVKNRYQFLNOLNPAIES 419
;

```



```

718 LLSIMDKTSTLPDVKNKSIITGANCQYEVNRINSEGRKELLASHGKWNKEAIMGDLSSK 777
776 EYISFNPKXENKIIVAKSKYHLSTLLQBIIRNANSSDIDLEKKVMLTCEINVASNIDRQ 835
778 EYIFDSIDNKLAKAKNIPGLASISEDIKTULLDASVSPDTKFILNNLKLNISSIGDY 837
836 IVEGRIBEAKNUTSDSINYIKNEFKLIESIDSLYDLKHONGLDSDSHFISPEDISKTEG 895
838 IYXEKLEPKYKNIHNSIDDLDEFNLLNVSDELYELKLANLDEKYLISFEDISKNNST 897
896 FPIRINKTGNISFIEETEKEFFSYVATHISKEISNIKDTTFDNNVNGKLVKKVNLDAHE 955
898 YSVRINKNGESVVETEKEFEFSKYSHEITKEISTIKNSIITDVGNNLLDNIQIDHTSQ 957
956 VNTLSAAPPFIQSLIBYNTTKESLNSLVAMKVQVYLAQLFSTGLNTITTDASKVVELVSTAL 1015
958 VNTLNAAPPFIQSLIDYSSNKDVLNDLSTSVKQVLAQLFSTGLNTIYDSIQLVNLISNAV 1017
1016 DETID 1020
1018 NDTIN 1022

RESULT 11
US-10-011-366-6
; Sequence 6, Application US/10011366
; Patent No. 6573003
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
; DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medien & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/011,366
; FILING DATE: 16-NO. 6573003-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,310
; FILING DATE: 23-OCT-1997
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 24-OCT-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-011-366-6

```

Query Match 53.8%; Score 2791; DB 4; Length 2710;
Best Local Similarity 52.1%; Pred. No. 4.6e-159;
Matches 534; Conservative 221; Mismatches 262; Indels 8; Gaps 5;

QY	1	MNLVNAQLOQKVVYKFRIOBDEVAITNALBESYHNMSESSVVEKYLKOLINNUDTNYL	60
Db	1	MSLISKEELIKLAY-SIRPRENEYKTLITLWDEYNKLTNNENYKYLQKLNESIDVFM	59
QY	61	NTYKSGRNKALKKFKPELWTVLEVKNSLTPVEKNLHFIWIGGOINDTAINYNOWD	120
Db	60	NYKTSRNRALSNLKDILKEVILIKNSVSPVEKNLHFVWIGVEVDIALEYIKQWAD	119
QY	121	VNSDYTVKVFYDSNAFLINTLKTIVESATNTLTSPRENLANDPDFYKNFYRKEMIY	180
Db	120	INABYNIKWLVDSEAFVWTLKKAIVBSSTEALQLLEBEIQNPQDNMFKYKMEFIY	179
QY	181	DKQKHFDYKSOJBENPEFIIDNIITKYLISNEYSKOLEALNKYIEESLNTKANNNDI	240
Db	180	DRQKFINYKSOJKNFVPTIDDIKSHLVSVNRDETVEJSTNSLRKINSNHGIDI	239
QY	241	RNLKFADEBVLVYNQELVERWNLAAASDILISMLEDGGVYLDVDILPQIODLFLKS	300
Db	240	RANSILFTQOELNYSQELLNRGNLAAASDIVLLALKNPFGGVYLDVMDLPGIHSDFKT	299
QY	301	INKPDSITNTGWEVIKLEAIMKYKEYIPGTYTSKNFMDLDEEVQSFESALSKSDKSEIF	360
Db	300	ISRPSSIGLRWEMIKLEAIMKYKYNNTSENFDKLDQOLKDNFKLIESKSEKSEIF	359
QY	361	LPDDIDKVSPLVKIAPANNVINOALISUKDSYCSDLVINQIKRYKILINDNLNPSINE	420
Db	360	SKLENLVSDIEIKIAFALGSVINQALISKQSGYLTNLVEIQKRYQFLQOHLNPALES	419
QY	421	GTDFNTMKIPSDKLASINEDNMFMKITNTLYKGFAPDVRSTINISGPGVYTGAYQD	480
Db	420	DNNFCTTKIIPHDSLFSNATAENSMTKTAPLYQVGFMPPEARSTISLSGPGVYASAYD	479
QY	481	LMWFKDNTNHLLEPELRNPFPPKTKISOLTQOETISLWSNQARAKSQSEYKKGVE	540
Db	480	FINLQENTIEKTRASDLIEFKFPENNLSLTQOETINSLWSFDQASAKYQPEKYVRYDTG	539
QY	541	GALGEDDNLDAQNTVLDKDY-VSKJLSS--WKTANKYIHYIVLOQDKDISVEASCNL	597
Db	540	GSLSEDNGVDFNKTADKNYLLNKNYPSNNVFEAGSKNYHYIYIQLQODDISYEATCNL	599
QY	598	FKSDYVSIYQKIEGSETAYYYVAD--AETKEIDKYRIPQISNKNIKLTFIGHGK	655
Db	600	FSKNPNBIIQRM--NESAKSVLSDGESILENKYRIPERLKNKVKVTFIGHGK	657
QY	656	SEFNTDTFANLDVDSLSEIETILNLAKADISPKYIEINLLGNMPSYSISAEETYPGKL	715
Db	658	DEFNTSEFARLSDVSLNESISFLDTIKLIDISPKVNEVNLGNMPSYDFNVEETYPGKL	717
QY	716	LKIKORYVELMPSISODSIYVSANOYEVEINEGKREILDHSGKWINKEESIIKDISKK	775
Db	718	LLSINMDKJTSTLPVKNKSTITIGANQVEAINSEGRKELLAHSGKWINKEBEAISDLSSK	777
QY	776	EYISFNPKENKIIKVS KYLHELSTLLOEIRNANNSDDILEKKVMLTECEINVASNIDRQ	835
Db	778	EYIFPDSIDNKLKAKSNIPGLASISBDIKTLILLDASVPDYKFINLNLKNISSIGDY	837
QY	836	IVEGRIEBAKNLTDSDINYIKNEFKLIESDSLYDLKHQNGLDOSHFTSFEDISKTENG	895
Db	838	IYVEKLEFPVKIINHSDILDEBNLLJENVSDELYELKLNLDKYLISFEDISKNKST	897
QY	896	FAIRIRKNTGNSPIETEKEIFSEYATHISKISINIKOTIPDNVGNKLKVKVNLDAHE	955
Db	898	YSVRFINNNGSVSVETEKEIFSKYSEHITKISITIKNSITIDVNGLLNDIQDHTSQ	957
QY	956	VNTLNSAFFIOLSHIEYNTKESLNSLVAMKVOVYAOFLSTGLNITDASKVVELJVSTAL	1015

Db 958 VNTLNRAFFIQSLDIYSSNKVDLNDLSTSVKQVQYAOLFSTGLNTIYDSQLVNLISNAV 1017

Qy 1016 DETD 1020

Db 1018 NDTIN 1022

RESULT 12

US-09-084-517-6

; Sequence 6, Application US/09084517

; Patent No. 6613329

; GENERAL INFORMATION:

; APPLICANT: KINK, JOHN A.

; APPLICANT: WILLIAMS, JAMES A.

; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND

; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL

; STREET: 220 MONTGOMERY STREET, SUITE 2200

; CITY: SAN FRANCISCO

; STATE: CALIFORNIA

; COUNTRY: UNITED STATES OF AMERICA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/084,517

; FILING DATE:

CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/
 FILING DATE: 16-MAR-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/329,154
 FILING DATE: 25-OCT-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/161,907
 FILING DATE: 02-DEC-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/985,321
 FILING DATE: 04-DEC-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/429,791
 FILING DATE: 31-OCT-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: CARROLL, PETER G.
 REGISTRATION NUMBER: 32,837
 REFERENCE/DOCKET NUMBER: OPFD-01610
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 705-8410
 TELEFAX: (415) 397-8338
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2710 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-084-517-6

	Query Match	53.8%;	Score 2791;	DB 4;	Length 2710;	
	Best Local Similarity	52.1%;	Pred. No. 4.6e-159;			
	Matches 534;	Conservative 221;	Mismatches 282;	Indels 8;	Gaps 5	
Qy	1	MLNVKAOLOQVYVKFRIODEYVAIINLAAEEYHNHSSSEWVEKYIKLKIDNNLTNYL	60			
		: : : : : : : : : : : : : : : : : :				
Dd	1	MSLIKEELIKLAY-SIRPRENEYKTITLDVEKLTTNNENKYLQLKKLNESIDVFM	59			
Ov	61	NTYKKSGRNKALKKFKEYLMEVLLEKNSLTIPVEKRLHFTIWGGINDTAINTYNOKWD	120			


```
Db 519 KVSQIINE-KSDDKRFPEDIKLARIISSELSIKGEKYSIKH-DGLDEFTTLNQLNIL- 575
Qy 352 SKSDSEIFLPLDDIKVSPLEVKIAFANNVI-----NQALISLKDSYC 395
Db 576 SEIEK---LITDDY-PKPVETKVVRDTFKIPKRYQKWTEWTNIRGNFMMLTHGSKC 630
Qy 396 SDLVINGIKRYKILNDLNPISNEGTFNTMKIFSDKLASISNE-----DNM 444
Db 631 IDFTLSGKKQYLEL-QRIRDNISYNLFTT-----EDLKSNNVAIGGIPAKKYLEHG 684
Qy 445 MFMIKITNYLKVGPADVRSTINLSGP----- 471
Db 685 LF-----SEYRQDGTIPYVSTLANISGPMIMRQMKYKISLGRIGEVIHKDKNKLSDVNFL 740
Qy 472 GVTGAGQDULLMF---KDSNNTHLLEPE-----LENFEPFKTSQLT---EQEITS 518
Db 741 GVIASSNKDNKSFNMLPVPVSGINDITPDDESSWAVRNNDINKILFEKINCHVPEKLPTS 800
Qy 519 LWSFNQARAKSQPEYKKGVPFEGALGEDDNLDFQAQNTVLDKDVSKKILSSMKTRNKEYI 578
Db 801 LY-----YEIDRSFPQGW---DNKSIXHVTIENDLI--KDNILLTSSNIDV 844
Qy 579 HYIVOLQDKISYBASCNLFSDKPYSSILYQKNEGSETAYYYVADAIEKIDKRIPI 638
Db 845 KLLIKL--DRELYAISSKI-----DNPLALRSIRTLQQLQANVYTSNTEPENTINFIY 896
Qy 639 QISNKN-----IKLFIHGKSEFNTDFANLDVDS-----LSSEIET 677
Db 897 DFYRKQDOLLSAIK-----FSRNDADTKIIVWNSYMEKNVFLREVISC 942
Qy 678 ILNLAKADISPKYTEINLGCNMFYSISAEETYPGKLLKIKDRVSELMPISQDSI-- 735
Db 943 VLRSKKVD---SYINEN-----KKNLSKEDA--GALRDYAKLKMELFSMLDDGYKK 990
Qy 736 TVSANQVEVRINEREGKEILDHSGKWINKBESIIKD:SSKEYISFNP-----K 783
Db 991 IITNAY--IKERDKL-----SGLIYNIENSIISGHESFDI:RSNOHEWGD:STVEQFK 1042
Qy 784 ENKIIVSKYLHEI:STLLOEIRNNANSSDIDLEKKVMLTECEINVAS-----NIDRQIVE 838
Db 1043 KFEFVYKSE-LSSAKIPDDIKNK-YITDPETKRNVLVHQLDSDIKERIAFLDISHVAYP 1100
Qy 839 GRIBEAKNLTSDSINYIKNEFKLIESI-----SDSLYDLKHQNGL 878
Db 1101 GSLLEKQLS---GYVFSDFINIIAEYLLASVGSVSHGVVYPAPSDKLELLRRHTK 1156
Qy 879 DSHFISPIDISKTENGPRIFINKETGNSI---FIETEKEIFSEYATHISKEISNIKT 935
Db 1157 SNSEWI--EKTP---YVVDILSDVSNVLRPPLSEBQKKILNDIKLEISKVS---EQ 1207
Qy 936 IFDNVNGKLVKKVNLDAAEVNTLNSAFFIQSLIEYNTTK-----ESTSNLSV----- 983
Db 1208 YFMKITEQKSSVIGIKYSVDFDRYNENLFLSLPINQNLTPFMYRYFEMLYDIIHIGIEN 1267
Qy 984 -AMKVQVVAQLFSTGLNITITDASKVEL 1010
Db 1268 KANREFIYKFSFSLNLDPLINDERVLNL 1295

RESULT 14
US-09-543-681A-4980
; Sequence 4980, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRTON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
```

```
; SEQ ID NO 4980
; LENGTH: 2807
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-4980

Query Match 5.1%; Score 267; DB 4; Length 2807;
Best Local Similarity 16.6%; Pred. No. 1.8e-07;
Matches 226; Conservative 217; Mismatches 432; Indels 488; Gaps 51;

Qy 2 NLNVKAQIQKVVYKFRIOEDEYVAILNALAEYHNMS---ESSVVEKVLKLDINNLTDN 58
Db 861 NLMTADELTSAEVTGARGENYKAIIOQLSDRYELOSTNTNTNTPIDIFQQJSTLRIOINS 920
Qy 59 YLNTYKSGRNKALKKPKKEYLTME-----VLEKXNSLTVPK 96
Db 921 YLLKHPDSCRNEALQQLRDQVDIIRYKHASILLSSQVDSNNTSYLYEISFANLKNK 980
Qy 97 NLHFWIGGQINDTAINYNQWKDVNSDYTVKVFYDSNAFLINTLTKKTIVE-----SATNN 152
Db 981 HIYL-----DENGNFVTKGN-NLQNOQDKLISGKVA--ITSIKRLVEKEYGSEITNQ 1029
Qy 153 TLESFRENLDPEFDYN-----KPYRKME-----IYDKQKH-----FI 187
Db 1030 VFNQFTEN---EPAQNGHGIDLSGLKKIHOAIEQOVSPISSTLYIWKPSKHSRLGHAAL 1085
Qy 188 DYKSKQIENPEFIIDNIKTYS-----NEYSKOLEA---LNKYIEE 227
Db 1086 QIGSGRLQDSDSVQDNHNNYVSWPAGSKSLANPLNITTEKNPDKLRWYDLSQPVSR 1145
Qy 228 SLN-----KITANNNDIRNLKFADELDVRLYNQELVERWNLAAASDI 271
Db 1146 STNFWTLQIDQEBEKTNFKLNKDEGENDRQJKFRD-----KTNLTRGLE- 1191
Qy 272 LRISMLEDGQVYLDVD-----ILPGIQDLPKSIKNPD-----SITNTSWEM 314
Db 1192 ---NITPEIAGLLANAPHWIPETLIPPEISHPFIKQWKS PNSNLLEVSTNFSIAVKWLK 1248
Qy 315 IKLEIMKYKEYIPGYTSKNPDMLEDEYQVSFESALSSKSKSEIF-LPLDDIKVSPLEV 373
Db 1249 VAPFTAMINQTP-ALEIKKIKEIEDRKTENPKTSVADDTNDGVVFRINLEGLDAAAMQ 1306
Qy 374 KIAPANNVINQALISLKDSYCSDLVINQIK----- 404
Db 1307 EWRKINSQ--SDPRYQLITKNCSSIVARILKAGAGADQLIGHHWKPRFGIWPTELYKFSQ 1364
Qy 405 -----NRYKILNDNLNPSINEGTDENTMTKIFSDK-L 435
Db 1365 KIQEARLAQIAVOQKPIHSLNQLALSDEHNKVAIDNDGTSP--NEKNLSPLTRFENDHFF 1422
Qy 436 ASISNEDNMFMIKITNYLKVGFAPDV-RST----- 465
Db 1423 GSVERRRDMTVMRIKNEEKITLKGAGRLTGSYYVRGNDNIPEATDKKVVLPFHGNSPT 1482
Qy 466 -----INLSGPGVYTGAYQDLMFKONSTNIHL-----EPB-- 497
Db 1483 EKQSSSFVHYNQOGIDMLAINMRGFGESDGSPTBQGMAYADAQTFRYLVDNDKGDPPKNI 1542
Qy 498 -LENFEFFPKISQLTE-----QETSLW-----SFNQARAKSQPEYKKGVPFEGALGED 546
Db 1543 IIHGYSMGAPIAAKLAASDISANGQHIAGLFLDRPMSMSKAIESYDEYSLKLTQLAKK 1602
Qy 547 DNLDEFA-----QNTVLDKQYVSK-----KILSSM 570
Db 1603 INGQFSVEKNLQFSKDIPIILLTDCGELGIGGEKWRTKLDDKGYLVKGERTDVSHLSL 1662
Qy 571 KTRNK-----EYIH----- 579
Db 1663 KLMNQYKQIISLSSLSPEYVNNTHRNALDSIQNPKNKVTDMGMYVYKPNNGGDSRF 1722
Qy 580 ----YIVOLQDKISYEASCNLFSDKPYSSILYQKNEGSETAYYYVADAIEKIDKRI 636
Db 1723 DSQIIITQEDDPIVAQAASALAAKAKNSIVVQLDADG-----HYRVA-----YGD 1768
```


637 PYQISNKENIKLTFIGHGK--SEFNTDTFANLVDLSLSSEI---ETILNLA-KADISPKY 690
 1769 PAQLSGK--IRWLVGHGRETTENNHLRLSNYNADLSNQLVKFNTVFSKNNINTTPEH 1826
 691 IEINLLGCMPSYSISABETPGKLLKIKDRVSELMPISQDSI--TVSANQVEYRINE 748
 1827 ISI--VGCSLISNDKQAGFAH-----QFIALDQOQIRSSVSARVTEVAVDS 1871
 749 EGKREILDHSGKWNKESIIKDISSEYISFNPKENKIIKSKYLHELSTLLQEIENNA 808
 1872 NGHYTKDNNGEWAQK-----NONKVLNWNKKGKITTEFQIHNGV 1914
 809 NSSDIDLEK-----KVLMECEINVASNIDRQIGVEIBEAKNLTSDSINYIKN---- 857
 1915 AEGDIDLTKIGASADNKVGAIDNNEVFTAPKKGKSTKSSGSDNSLSYSNGIQVS 1974
 858 -----BFLKIESDSL-----YDLKHQGLDDSHFISFED 888
 1975 VGDGEFTTVNWGTSNLGIKVGFGKSLVFGDNNVWVHVGNGDSKH--SVDIAGYQAFEG 2032
 889 ISKTENGFRIRINKETGNSFIETKEIPS-----EYATHISKEISNIKDTIFDNVN 941
 2033 MQLFVGTNRVNF--NLGRSNDLWMLEKSIPTPLINPFGGAARITVKLESACSGENNDN 2091
 942 GKX-----VKKVNLDAAEHVNTLNSAFFIQSLIEYNT 973
 2092 DMLSVQNDWTLGAKKFAIDMSGIDQTSN-----VDYKT 2126

RESULT 15
 US-09-417-485D-6
 ; Sequence 6, Application US/09417485D
 ; Patent No. 6541202
 ; GENERAL INFORMATION:
 ; APPLICANT: Long, David M.
 ; APPLICANT: Metz, Anneke M.
 ; APPLICANT: Love, Ruschelle A.
 ; TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes
 ; FILE REFERENCE: 47714-5009-US
 ; CURRENT APPLICATION NUMBER: US/09/417,485D
 ; CURRENT FILING DATE: 2002-06-14
 ; NUMBER OF SEQ ID NOS: 49
 ; SOFTWARE: Patentn Ver. 2.1
 ; SEQ ID NO 6
 ; LENGTH: 2184
 ; TYPE: PRT
 ; ORGANISM: Plasmodium falciparum
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (330)..(335)
 ; OTHER INFORMATION: Xaa at position 330 = Leu or Ile;
 ; OTHER INFORMATION: Xaa at position 335 = Asp or Gly.
 US-09-417-485D-6

Query Match 5.0%; Score 257.5; DB 4; Length 2184;
 Best Local Similarity 19.6%; Pred. No. 4.7e-07;
 Matches 251; Conservative 192; Mismatches 446; Indels 391; Gaps 60;
 QY 10 QKVVYKFRQEDBYVAILNALESEY--HAMSES---SVVEKYLKLDI-----NNLTD 57
 572 KKCIPKLLGSKNFKIFLQVKKFLLENFYKESFSLNQVKNKIKVKNIFQKKISKYIKN 631
 58 NYL--NTYKSGRNKAL--KKFKEVLT-----MEVLEKNSLTPPEKNLHFIWIGGQIND 109
 632 RILLKNIFDNNYENKILHRNKKIITINDNIIKYNKQDNL-----NNSF-----KIKT 681
 110 TAIN-----YINQWQVNSDYTVK-----VFVDSNAFLINTLUKTVIVESATNNTL---- 154
 682 TLENKLRKYFNKIKINIAIQKHLNRLIYLFVNFIMPLIRPFPLTKSEQTLHKTI 741
 155 -----ESPR-----ENLNDPEFDY--NKP-YKGMELIYDKQKHEI 187

Db 742 FDRKIKWNHFTKISNFCULYHQIFRNKCLKKRNBPMDYVQNMFNVKKKGEKI--KTNKYI 799
 QY 188 DYKQOIENPEFIIDNIITKYLSEYKDJLEAL-----NKYIEBSLANKITANN--GNDI 240
 Db 800 FIKQKKSTKINCINNPKSCIKPIKCKKQNLN--TRNNNIFIKKMEKSKTNNLINKSI 859
 QY 241 RNLEFAD--EDLVLYNQEL---VERNLAASAIIIRISMLKEDGGVYLDVDILPQIP 295
 Db 860 DNLKELKEINKSVAPYKFKYIKKKYFALKWYIHRMAKEE----- 904
 QY 296 DLFKSIKPDSTINTSWEMIKLEAIMKY-----KEYIPGYTS-----KNFD 336
 Db 905 ---KS-----NIKLERAFKHPFFIAQKEHILKYFSSHFFQNRKNKYKGFN 948
 QY 337 MLDDEVQ-----RSPESALSSKSDKSEIFLPLDDIKVSPLEVKIAFANNS 381
 Db 949 KLHRIKIIIIKQNGSIVKQKDKTFLHLTKNKNKN-----NNK 988
 QY 382 VINQALISKDSYCDLVINQIKRYKILNDNLNLSINEGTDFNTMTKIFSDKLASIGNE 441
 Db 989 KKN-----KNYNN-----NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1033
 QY 442 DNMFMIKITNLYKVGFPADVRSTINLSPGVYTGAYQDLL-----MFKDSTNIHLEP 496
 Db 1034 NNNK-----KAKNEKNIDD--SNLEKKKKIYIKKIIIEKRNFMKLKNSIN-HFISK 1086
 QY 497 ELRNFEFPKTK-----ISQTEQBITSLMFOQARAKSQPEEYKKGFFEGALGEDDML 549
 Db 1087 KLRINWPKKGLRPLINLSTLNVPEIVKQRIFFELKSKSSE---FYFHNILANLERE 1142
 QY 550 DPAQNTVLQDKYVSK-----KILSMKTPN-----KE 576
 Db 1143 KQKNIKQKRYKKNFNPNVSLNINCNFSLKCLGNMHRNNSLFXNTLTKTGEIEBKALK 1202
 QY 577 YIHVIV-----QLOGDKISYEASC-----NLFS--KDPYS 605
 Db 1203 WELHYKQWFKYKRMKKYIKNKKNNKKIYAYICIGDFNCYEHINHNLYFLKLNFFDN 1262
 QY 606 I-----LYQKNIESTAYY-----YVADAEIKEIDKVRIPVOISN 642
 Db 1263 INNFEFYLKRSFRLYNKNLNSLSLYPVNVKSGLHY-----IRNLRELIIKSLAND 1317
 QY 643 KRNIKLTFIGHGKSEFN-----TDTFANLVD--SLSEIETILNLAKAISPKYIENLL 696
 Db 1318 NHHFLLNQMEKTKSDLYIFADSYKSLQVDRKIDIFMTIIVIRYYLYNIYFSIKEFKLN 1377
 QY 697 GCMFYSISAEETYPQKLL-LAIKQV-----SELMPSISQDSITVSA 739
 Db 1378 RKNIFYQIQENQMGVLSVDRKRVENIKKWLNSMKKINHDELSLKNSSININ 1437
 QY 740 NQYEVRIINEBKGEILDHSGKWNKESIIK-----DISKKEYISFNPKENKIYK 790
 Db 1438 KNFMICTNHEQDTE---EKGNTQKKEKHDYIGPIYNNSPDSTTTTHSSNNYKNNIHVS 1494
 QY 791 SKY-----LHELSTLQF-----IRNNANSDDIDLEKKVMLTECEINVASNIDRQIVE 838
 Db 1495 GDYKNDGLLHKGNNSMNECYVKIKCNKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1553
 QY 839 GRIBEAKNLSDSI---NYIKNEFK--LIESISDSLYDLKHQGLDD---SHRISFEDIS 890
 Db 1554 --IKYHTIDTDSKNHTYFNKFLNPLDKKIIISNIYGLPQGFSLNLSLYAYLD-- 1609
 QY 891 KTEGFRIRINKETGNSFIETKEIFSEVATHISKEISNIKDTIFDNVNGK----- 943
 Db 1610 KNEEFQNLLEYSEKQINNKYFLANGT---CNYFNLSLILRFDLFLITLKNKIKFKN 1666
 QY 944 --LVKVNLDAAHVBWNTLNSAFFIQSLI-----EYNTTKESLSNLSVAMKV-- 987
 Db 1667 LLLKKKI---WGSNINSKTKIFKIPLIYKNDLLIYNFQNKYQKKYKIKKKKIQSVRN 1723
 QY 988 -QVYAQLFSTGLNTITDASK 1006
 Db 1724 KEIHQLVWANKKHTSVQK 1743

FFI Apr 7 08:15:50

Search completed: April 1, 2004, 16:45:56
Job time : 23.375 secs

CC protein, are used to raise avian antibodies useful as antitoxins or
CC diagnostics, and in vaccine prodn. (Updated on 16-OCT-2003 to standardise
CC OS field)
XX
SQ Sequence 2366 AA;

Query Match 79.0%; Score 4101; DB 2; Length 2366;
Best Local Similarity 77.6%; Pred. No. 1.4e-213;
Matches 792; Conservative 106; Mismatches 122; Indels 0; Gaps 0;
QY 1 MNLVNAQLOKQVYKPIQDEYVAIILAEVEYHNHNSSESVVEKYLKLDINLTDNYL 60
DB 1 MSLVNRKQLEKMANVRFTQDEYVAIILAEVEYHNHNSSESVVEKYLKLDINSLTDIYI 60
QY 61 NTYKSGRNKALKKFKFKEVLTMEVLELKNNSLTPVEKNLHFVIGQINDTAINYNQWKD 120
DB 61 DTYKSGRNKALKKFKFKEVLTMEVLELKNNSLTPVEKNLHFVIGQINDTAINYNQWKD 120
QY 121 VNSDYTVKVFYDSNAFLNTLTKTIVESATNTLIESFRENLDNPFDPYKFRMEIY 180
DB 121 VNSDYTVKVFYDSNAFLNTLTKTIVESATNTLIESFRENLDNPFDPYKFRMEIY 180
QY 181 DQKHFDYKQISENPEFIIDNIITKYLSEYKQLEALNKYIEESLNKITANNNDI 240
DB 181 DQKHFDYKQISENPEFIIDNIITKYLSEYKQLEALNKYIEESLNKITANNNDI 240
QY 241 RNLKFADEDLVRLNQELVERNLAAASDLIRISMLKEDGGVLDVLDLPGIQDLPFS 300
DB 241 RNLKFADEDLVRLNQELVERNLAAASDLIRISMLKEDGGVLDVLDLPGIQDLPFS 300
QY 301 INKPSITNTSWMIKLEAIMKYKPIGYTSKNFMDLDEEVQSFESALSKSKDKSEIF 360
DB 301 INKPSITNTSWMIKLEAIMKYKPIGYTSKNFMDLDEEVQSFESALSKSKDKSEIF 360
QY 361 LPDDIKVPLEVKAFANNSVINQALISLSDSYCSGLVINOIKNRYKILANDNPSNE 420
DB 361 LPDDIKVPLEVKAFANNSVINQALISLSDSYCSGLVINOIKNRYKILANDNPSNE 420
QY 421 GTDENMTKIFSKLASINENMMFMKIITNLYKVGAPDVRSTINLSGPOVYTGAYQD 480
DB 421 GTDENMTKIFSKLASINENMMFMKIITNLYKVGAPDVRSTINLSGPOVYTGAYQD 480
QY 481 LLMFKDNSTNIHLEPELNFPPFKTKISQLETSITSLWSFNQARAKSQFEYKKGVE 540
DB 481 LLMFKDNSTNIHLEPELNFPPFKTKISQLETSITSLWSFNQARAKSQFEYKKGVE 540
QY 541 GALGEDNLDPAQNTVLDKDYVSKKTLSSKTRNKEYIHYIVQLQDKISYASCNLSK 600
DB 541 GALGEDNLDPAQNTVLDKDYVSKKTLSSKTRNKEYIHYIVQLQDKISYASCNLSK 600
QY 601 DPYSSILYQKNIEGSTAYVYVADAEIKELDKYRTPYQISKNKIKLTFIGHGKSEFT 660
DB 601 DPYSSILYQKNIEGSTAYVYVADAEIKELDKYRTPYQISKNKIKLTFIGHGKSEFT 660
QY 661 DTPANLDVDSLSSEITILNLAADISPKYIEINLGCNMFYSISAEETYPKLLLIK 720
DB 661 DTPANLDVDSLSSEITILNLAADISPKYIEINLGCNMFYSISAEETYPKLLLIK 720
QY 721 DRYSELMPISQDSITVSANQYEVRIINBEKREILDSGKWNKESIIKDISKEYISF 780
DB 721 DRYSELMPISQDSITVSANQYEVRIINBEKREILDSGKWNKESIIKDISKEYISF 780
QY 781 NPENKTIIVSKYVHELSTLQIRNANSSDIDLEKKVMTCEINVASNIDRQIVEGR 840
DB 781 NPENKTIIVSKYVHELSTLQIRNANSSDIDLEKKVMTCEINVASNIDRQIVEGR 840
QY 841 IEEAKNLTSINIKNEPKLIESISDLYDLKHONGLDSDHIFSPEDISCKENGFIRF 900
DB 841 IEEAKNLTSINIKNEPKLIESISDLYDLKHONGLDSDHIFSPEDISCKENGFIRF 900
QY 901 INKETGNSFIETEKETFEYSYATHISKEISNIKDTIFDNYNGKLVKKNVLDAAHEVNTLN 960
DB 901 INKETGNSFIETEKETFEYSYATHISKEISNIKDTIFDNYNGKLVKKNVLDAAHEVNTLN 960

QY 961 SAFFIQSLIEYNTTKESLSNLSVAMKVOVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
DB 961 SAFFIQSLIEYNTTKESLSNLSVAMKVOVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
RESULT 2
AAW68388
ID AAW68388 standard; protein; 2366 AA.
XX
AC AAW68388;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium difficile toxin B.
XX
KW Antitoxin; vaccine; cytotoxin; toxin B; intoxication; immunogen;
KW pseudomembranous enterocolitis.
XX
OS Clostridium difficile.
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US015394.
XX
PR 28-AUG-1996; 96US-00704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Williams JA, Thalley BS;
XX
DR N-PSDB; AAV30561.
XX
PT Host cell containing recombinant expression vector encoding Clostridium
PT botulinum type B or E toxin - useful to treat humans and other animals at
PT risk of intoxication with clostridial toxin.
XX
PS Example 18; Page 241-249; 428pp; English.
XX
CC This is the amino acid sequence of Clostridium difficile toxin B, deduced
CC from the coding region (see AAV30561) of the toxin B gene. Fragments of
CC the toxin B gene have been cloned into various prokaryotic expression
CC systems, and assessed for the ability to express recombinant toxin B
CC protein in E. coli. It would be advantageous to use simple and
CC inexpensive prokaryotic expression systems to produce and purify high
CC levels of recombinant toxin B for immunisation purposes. The invention
CC specifically relates to recombinant proteins derived from Clostridium
CC botulinum toxins (see AAW68389-400) and their use as immunogens for the
CC production of vaccines and antitoxins
XX
SQ Sequence 2366 AA;
Query Match 79.0%; Score 4098; DB 2; Length 2366;
Best Local Similarity 77.5%; Pred. No. 2e-213;
Matches 791; Conservative 107; Mismatches 122; Indels 0; Gaps 0;
QY 1 MNLVNAQLOKQVYKPIQDEYVAIILAEVEYHNHNSSESVVEKYLKLDINLTDNYL 60
DB 1 MSLVNRKQLEKMANVRFTQDEYVAIILAEVEYHNHNSSESVVEKYLKLDINSLTDIYI 60
QY 61 NTYKSGRNKALKKFKFKEVLTMEVLELKNNSLTPVEKNLHFVIGQINDTAINYNQWKD 120
DB 61 DTYKSGRNKALKKFKFKEVLTMEVLELKNNSLTPVEKNLHFVIGQINDTAINYNQWKD 120
QY 121 VNSDYTVKVFYDSNAFLNTLTKTIVESATNTLIESFRENLDNPFDPYKFRMEIY 180
DB 121 VNSDYTVKVFYDSNAFLNTLTKTIVESATNTLIESFRENLDNPFDPYKFRMEIY 180
QY 181 DQKHFDYKQISENPEFIIDNIITKYLSEYKQLEALNKYIEESLNKITANNNDI 240
DB 181 DQKHFDYKQISENPEFIIDNIITKYLSEYKQLEALNKYIEESLNKITANNNDI 240

Db 181 DKQKFNINYYKAQRENPELLIDIVKTVLSNEYSKEIDELNTYIIEESLNKLTQNSGNDV 240
 Qy 241 RNLEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVDPILGQIDPLFKS 300
 Db 241 RNLEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVDPILGQIDPLFKS 300
 Qy 301 INKPSDITNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEEVQVSFESALSKSDKSEIF 360
 Db 301 IEKPSVTVDFWEMTKLEAIMKYKEYIPGYTSKNFDMLEEVQVSFESALSKSDKSEIF 360
 Qy 361 LPDDDKVSPLEVKIAPANNVINQALISLKDSCDVLVINOIKRYKILNOLNAPSINE 420
 Db 361 SSLGDMEASPLEVKIAPANNVINQALISLKDSCDVLVINOIKRYKILNOLNAPSINE 420
 Qy 421 GTDFTNTMTKIFSDKLASISNEDNMFMKIITNLYKVGFPADVRSTINLSGPGVYTGAYOD 480
 Db 421 DNDFTNTMTKIFSDKLASISNEDNMFMKIITNLYKVGFPADVRSTINLSGPGVYTGAYOD 480
 Qy 481 LLMFKDNSTNIHLLPELRNFFPKTKISQLETEQITSLWSFNQARAKQFYEYKGYE 540
 Db 481 LLMFKDNSTNIHLLPELRNFFPKTKISQLETEQITSLWSFNQARAKQFYEYKGYE 540
 Qy 541 GALGEDNLDPQAQTVLDKDYVSKLILSSMKTNRKEYIHYIVOLQDKISYEASCNLFSK 600
 Db 541 GSGEDDNLDPQAQTVLDKDYVSKLILSSMKTNRKEYIHYIVOLQDKISYEASCNLFSK 600
 Qy 601 DPYSILYOKNIEGSETAYVYVADAEIKIDKVRIPQISNENIKLTFIGHGKEFNT 660
 Db 601 TPYDVLFOKNIEGSETAYVYVADAEIKIDKVRIPQISNENIKLTFIGHGKEFNT 660
 Qy 661 DTFANLVDLSSEIETILNADISPKVIEINLGCNMFYSISAEETYPGKLILKIK 720
 Db 661 DTFANLVDLSSEIETILNADISPKVIEINLGCNMFYSISAEETYPGKLILKIK 720
 Qy 721 DRVSELMPSISQDSITVSANQYVRINEEGKREILDHSGKWINKESIIKDSSKEYISF 780
 Db 721 DRVSELMPSISQDSITVSANQYVRINEEGKREILDHSGKWINKESIIKDSSKEYISF 780
 Qy 781 NPKENKIIVKSKYHLSTLLOETRNANSSDIDLEKVMTECEINVASNIDQIUEGR 840
 Db 781 NPKENKIIVKSKYHLSTLLOETRNANSSDIDLEKVMTECEINVASNIDQIUEGR 840
 Qy 841 IEAKNTLSDSINIKYKFEKLIBSISDLVDLKHONGLDSDHFSFEDISKTENGFRIF 900
 Db 841 IEAKNTLSDSINIKYKFEKLIBSISDLVDLKHONGLDSDHFSFEDISKTENGFRIF 900
 Qy 901 INKGTGSIETETEKEIFSEYATHISKEISNIKDTIFDNVNGKLVRKVNLDAAHEVNTLN 960
 Db 901 INKGTGSIETETEKEIFSEYATHISKEISNIKDTIFDNVNGKLVRKVNLDAAHEVNTLN 960
 Qy 961 SAFFIQSLIEYNTTKESLSNLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
 Db 961 SAFFIQSLIEYNTTKESLSNLVAMKVQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
 RESULT 3
 ID AAY33700 standard; protein; 546 AA.
 XX AAY33700;
 AC AAY33700;
 XX 18-JAN-2000 (first entry)
 DT 18-JAN-2000 (first entry)
 DE C. sordellii lethal toxin protein fragment.
 XX Lethal toxin; immunotoxin; antitumor; Glucosyl transferase; glycosylate;
 KW GTPase; Ras; epidermal growth factor stimulated MAP-kinase; inhibition;
 KW signalling pathway; cell-specific toxin; treatment; cancer.
 OS Clostridium sordellii.
 XX DE19802569-A1.
 PN DE19802569-A1.
 XX

PD 09-SEP-1999.
 XX 23-JAN-1998; 98DE-01002569.
 XX 23-JAN-1998; 98DE-01002569.
 XX (UYFR-) UNIV FREIBURG ALBERT-LUDWIGS.
 XX Aktories K, Hofmann F;
 WPI; 1999-509323/43.
 DR N-PSDB; AAZ23800.
 XX New fragment of the lethal toxin from Clostridium bacterium, useful for
 PT treating cancer.
 PS Claim 1; Page 7-9; 14pp; German.
 XX This invention describes a novel fragment (I) of the lethal toxin (LT) of
 CC Clostridium sordellii which has antitumor activity. (I) is a Glucosyl
 CC transferase that glycosylates, and thus inactivates, GTP(guanine
 CC triphosphate)ases, particularly Ras (an oncogenic product overexpressed
 CC in many tumors), resulting in inhibition of epidermal growth factor
 CC stimulated MAP-kinase signalling pathways. (I), particularly in the form
 CC of immunotoxins, are used as cell-specific toxins, particularly for
 CC treating cancer. When included in immunotoxins, (I) can be targeted to
 CC selected cells. Compared with the complete LT, (I) is smaller, so enters
 CC cells more easily, resulting in greater toxicity in the cytosol, is less
 CC likely to induce formation of (neutralizing) antibodies, and is more
 CC active than the holotoxin. This sequence represents the lethal toxin
 CC fragment described in the invention
 XX
 SQ Sequence 546 AA;
 Query Match 53.8%; Score 2792; DB 2; Length 546;
 Best Local Similarity 99.5%; Pred. No. 2.e-143;
 Matches 543; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 MNLVNAQLQKVVYKFRIOEDYVAILNLEEEYHNMSESSVVEKYLKLDKINLTDNYL 60
 Db 1 MNLVNAQLQKVVYKFRIOEDYVAILNLEEEYHNMSESSVVEKYLKLDKINLTDNYL 60
 Qy 61 NTYKSGRNKALKKPKFYLTVELKNNSLTPVEKNLHFHWIGQINDTAINYNQWKD 120
 Db 61 NTYKSGRNKALKKPKFYLTVELKNNSLTPVEKNLHFHWIGQINDTAINYNQWKD 120
 Qy 121 VNSDYTVKVFYDSNAFLINTLKTIVESATNTLTSPRENLDNPFYKFKRKEWIIY 180
 Db 121 VNSDYTVKVFYDSNAFLINTLKTIVESATNTLTSPRENLDNPFYKFKRKEWIIY 180
 Qy 181 DKQKHFIDYKSOIEENPEFIIDNIIKTVLSNEYSKDLKALNKYIEESLNKLTANNNDI 240
 Db 181 DKQKHFIDYKSOIEENPEFIIDNIIKTVLSNEYSKDLKALNKYIEESLNKLTANNNDI 240
 Qy 241 RNLEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVDPILGQIDPLFKS 300
 Db 241 RNLEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVDPILGQIDPLFKS 300
 Qy 301 INKPSDITNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEEVQVSFESALSKSDKSEIF 360
 Db 301 INKPSDITNTSWEMIKLEAIMKYKEYIPGYTSKNFDMLEEVQVSFESALSKSDKSEIF 360
 Qy 361 LPDDDKVSPLEVKIAPANNVINQALISLKDSCDVLVINOIKRYKILNOLNAPSINE 420
 Db 361 LPDDDKVSPLEVKIAPANNVINQALISLKDSCDVLVINOIKRYKILNOLNAPSINE 420
 Qy 421 GTDFTNTMTKIFSDKLASISNEDNMFMKIITNLYKVGFPADVRSTINLSGPGVYTGAYOD 480
 Db 421 GTDFTNTMTKIFSDKLASISNEDNMFMKIITNLYKVGFPADVRSTINLSGPGVYTGAYOD 480
 Qy 481 LLMFKDNSTNIHLLPELRNFFPKTKISQLETEQITSLWSFNQARAKQFYEYKGYE 540
 Db 481 LLMFKDNSTNIHLLPELRNFFPKTKISQLETEQITSLWSFNQARAKQFYEYKGYE 540

QY	121	VNSDYTVKVFYDSNAFLINTLTKTIVESATNTLTSPRENLDPEFDYKFKRMELIY	180
Db	120	INAEYNIKLWYDSEAFVNTLKAIVESSTTALQLLEEIQNFQDNMKFKRMELIY	179
QY	181	DKQKHFIDYKSOIEENPEFIIDNIITKLYLSNEYSKDEALNKYIEESLNKITANNGNDI	240
Db	180	DRQKGFINYKSOINKPTVPTIDDIILKSHLVSENRDETIVLESYRTNLRKINSNHGDI	239
QY	241	RNLEKFADEDLVRLYNQBELVERWNLAASDILRLISMLKEDGCVYLDVLDILPGIOPDLFKS	300
Db	240	RANSLFTEQELLNIYSQELLNRGNLAASDIVRLALAKNFGVYLDVDMVLFQIHSDLFKT	299
QY	301	INKPDSITNTSWEMIKLEAIMKYEYIPYTSKPFMDLDEEVQSFESGALSCKSKXSIF	360
Db	300	ISRPSSIGLDRWEMIKLEAIMKYKYINNYTSENFDKLDQQLKONFKLIIESKSEKSEIF	359
QY	361	LPDDIKVSPLEVKJAFANNVINOALISLKSYSVDIVINQIKNRYKILANDLNPSINE	420
Db	360	SKLENLVNDEIKIAFALGVSINOALISKOGSYLTNIVIEQVKRYQFLAHLNPAYES	419
QY	421	GTDENTTKIPSDKLASINEDNMFMKITNLYKVGFAPOVRSTINLSGPGVYTGAYQD	480
Db	420	DNNFTDTTKIFHDSLFNSATAENSFLTKIAPYLOVGFMPPEARSTISLSGPGAYASAYD	479
QY	481	LLMFKDNSTNIHLPELRNFPFKTKISQLTEQBITSLSFNQARAKSQPEYKKGYFE	540
Db	480	FINLQENTIEKTLKASDLIEFFKPPENNLSQLTEQINSLSWSPDQASAKYQPEKYVRYDTG	539
QY	541	GALGEDNDLDFACNTVLDKDY-VSKKLSS--MKTENKEYIHYIYVLOQDKLSYEASCNL	597
Db	540	GSLSEDNGVDFNKNTALDKRYLLNNKIPSNVVEAGSKNYVHYIYVLOQDDLSYEATCNL	599
QY	598	FSKDPYSSILYQKNTGSETAYVYVAD--ABEIKEDIKRYIPYQISNENIKLTFIGHGK	655
Db	600	FSKNPKNSIIQRNM--NESAKSYFLSDDGESILELNKYRIPERLNKEKVKVTFIGHGK	657
QY	656	SEFNTDTFRANLDVDSLSSSEIETILNLAKADISPKYIEINLLGCNMFSYSISAEZTPGKL	715
Db	658	DEFTSEFARLSVDSLSSNISSEFLDTIKLIDISPKVVEVNLGCNMFSYDFNVEETYPGKL	717
QY	716	LLKIKDRVSELMPFSIQSDSITVSANQYEVRIINEEGREILDSHGKWKINKESIIKDISSK	775
Db	718	LLSIMDKITSLPDVNNKNSITIGANQYEVRIINSEGRKELLASHGKWKINKEEAIMSDLSK	777
QY	776	EYISFNPKNKIIVKSKYLHELSTLLOEIRNNANSSDIDLEKKVMLECEINVASNIDRQ	835
Db	778	EYIFFSDIDNKLKAKSKNIPGLASISEDIKTLILDASVSPDTKFIILNNLKMIIESSIGDY	837
QY	836	IVEGRIEEAKNLTSDSINVIKNEFKLIESISDLYDLKHQNGLDSDHSFISFEDISKTENG	895
Db	838	IYYEKLPEPVKNIITHNSIDDLIDFNLENNVSDLEYELKLNLDLDEKYLISFEDISKNST	897
QY	896	FRIRFINKEGNSIFITEKEIPSEYATHISKEISNIKDTIFDNVNGKLVKKNVLDAAHE	955
Db	898	YSVRFINKNESGVYVETKEIPSKYSEHITKEISTIKNSIITDVNGNLLDNQLDHTSQ	957
QY	956	VNTLNSAFFTQSLIEYNTTKESLSNLSVAMKQVVAQLFSTGLNTITDASKVVELVSTAL	1015
Db	958	VNTLNAAFFIQSLIDYSSNKNVLDLSTSVKQVLAQLFSTGLNTIYDSIQVLNLSNAV	1017
QY	1016	DETID 1020	
Db	1018	NDTIN 1022	
RESULT 5			
AAW68387			
ID	AAW68387	standard; protein; 2710 AA.	
XX	AC	AAW68387;	
XX	DT	07-DEC-1998 (first entry)	
XX	XX		

QY	541	GALGED 546	QY	541	GALGED 546
Db	541	GALGED 546	Db	541	GALGED 546
RESULT 4					
AAR95016					
ID	AAR95016	standard; protein; 2710 AA.			
XX	AC	AAR95016;			
XX	DT	16-OCT-2003 (revised)			
XX	DT	08-JUL-1996 (first entry)			
XX	DE	C. difficile toxin A.			
XX	KW	Toxin A; cytotoxin; enterotoxin; fusion protein; antitoxin; vaccine;			
XX	KW	diarrhoea; diagnosis; therapy.			
XX	OS	Clostridium difficile; VPI strain 10463 (ATCC 10463).			
XX	PN	W09612802-A1.			
XX	PD	02-MAY-1996.			
XX	PF	23-OCT-1995; 95WO-US013737.			
XX	PR	24-OCT-1994; 94US-00329154.			
XX	PR	16-MAR-1995; 95US-00405496.			
XX	PR	14-APR-1995; 95US-00422711.			
XX	PR	07-JUN-1995; 95US-00480604.			
XX	PA	(OPHI-) OPHIDIAN PHARM INC.			
XX	PI	Williams JA, Padhye NV, Kink JA, Thalley BS, Stafford DC;			
XX	PI	Firca JR;			
XX	DR	WPI; 1996-230603/23.			
XX	DR	N-PSDB; AAT29248.			
XX	PT	Fusion proteins comprising non-toxin protein and part of toxin - useful			
XX	PT	to form anti-toxins against Clostridium botulinum type A, and C.			
XX	PT	difficile type toxins, and to treat C. difficile intoxication, partic.			
XX	PT	diarrhoea.			
XX	PS	Claim 63; Page 290-302; 434pp; English.			
XX	CC	Clostridium difficile VPI strain 10463 toxin A (AAR95016), the product of			
XX	CC	the toxin A gene (AAT29248), is a potent cytotoxin that plays a direct			
XX	CC	role in damaging gastrointestinal tissues and is associated with			
CC	CC	diarrhoeic disease. It can be obtd. by expression in transformed E. coli			
CC	CC	hosts of portions of DNA that together cover the entire toxin A gene.			
CC	CC	Toxin A, and portions of it (see also AAR95014-15 and AAR95017), pref.			
CC	CC	expressed as fusions to polyhistidine affinity tags or maltose binding			
CC	CC	protein, are used to raise avian antibodies useful as antitoxins or			
CC	CC	diagnostics, and in vaccine prodn. (Updated on 16-OCT-2003 to standardise			
CC	CC	OS field)			
XX	XX	Sequence 2710 AA;			
XX	XX	Query Match 53.8%; Score 2791; DB 2; Length 2710;			
XX	XX	Best Local Similarity 52.1%; Pred. No. 2e-142;			
XX	XX	Matches 534; Conservative 221; Mismatches 262; Indels 8; Gaps 5;			
QY	1	MNLVNAQLOQVYKFRIOEBEYVAILNALNEEYHNMSSSVVEKYLKDKINNTDNYL 60			
Db	1	MSLISKEELIKLAY-SIRPRENEYKTLTNLDYNNKLTNNENKYLQKLKLNESIDVFM 59			
QY	61	NTYKSGRVALKFKFEYLTMEVLEKNSLTPVEKNLHFIMTGGQINDTAINYNQMD 120			
Db	60	NKYTSRNRALSNNLKDKILKEVILKNSTSPVEKNLHFVWIGVEVSDIALEYIKQMD 119			

DE	Clostridium difficile toxin A.	
XX	Antitoxin; vaccine; cytotoxin; toxin A; intoxication; immunogen;	
KW	pseudomembranous enterocolitis.	
XX		
XX	Clostridium difficile.	
XX		
PN	W09808540-A1.	
XX		
PD	05-MAR-1998.	
XX		
XX	28-AUG-1997; 97WO-US015394.	
PF		
XX		
XX	28-AUG-1996; 96US-00704159.	
PR		
XX	(OPHI-) OPHIDIAN PHARM INC.	
XX		
XX	Williams JA, Thalley BS;	
PI		
XX	WPI; 1998-230234/20.	
DR	N-PSDB; AAV30560.	
DR		
XX		
XX	Host cell containing recombinant expression vector encoding Clostridium	
PT	botulinum type B or E toxin - useful to treat humans and other animals at	
PT	risk of intoxication with clostridial toxin.	
XX		
PS	Example 15; Page 220-230; 428pp; English.	
XX		
CC	This is the amino acid sequence of Clostridium difficile toxin A, deduced	
CC	from the coding region (see AAV30560) of the toxin A gene. Toxin A is a	
CC	potent cytotoxin that plays a direct role in damaging gastrointestinal	
CC	tissues. Severe cases of C. difficile intoxication result in	
CC	pseudomembranous colitis. This would be prevented by neutralising the	
CC	effects of toxin A in the gastrointestinal tract. Examples are provided	
CC	of the production of recombinant C. difficile toxin A in host cells and	
CC	of the in vivo neutralisation of toxin A by antibodies against	
CC	recombinant toxin A polypeptides. The invention specifically relates to	
CC	recombinant proteins derived from Clostridium botulinum toxins (see	
CC	AAW6389-400) and their use as immunogens for the production of vaccines	
CC	and antitoxins	
XX		
SQ	Sequence 2710 AA;	
	Query Match 53.8%; Score 2791; DB 2; Length 2710;	
	Best Local Similarity 52.1%; Pred. No. 2e-142;	
	Matches 534; Conservative 221; Mismatches 262; Indels 8; Gaps 5;	
Qy	1 MNLVKAQLQGVTVKFRIDEYVAIINALEEYHNMSSESVVEKYLKXDINNLDNYL 60	
Db	1 MSLISKELIKLAY-SIRPRENEVKITLNLDEINKLTNNENKYLQKLKLNESIDVFM 59	
Qy	61 NTYKSGRNKALKPKFELYTMVEVLEKNNSLTPVERKLNHFIIWGGQINDPAINYNQWKD 120	
Db	60 NKYKTSRRNALSNLKXDKILKEVILIKNSNTSPVEKNLHFVWIGGEVSDIALEVIKQWAD 119	
Qy	121 VNSDYTVKVDNAPLINTLKTYIVESAATNTLESFRENLDPEFPYKFKREIILY 190	
Db	120 INAEYNIKLWYDSEAFVNTLKKAIVESSTEAQLLEBEEIQNPQPDNMFKYKGRMEFIY 179	
Qy	181 DKQHFIDYKKSQJEEENPEFIIDNIITKYLNSNEYSKDLALNKYIBESLNKTIANNNGDI 240	
Db	180 DRQKRFINYKSKQNKPTVPTIDIIKSHLVSEYNRDETLESVRTNLSLRKINSNGIDI 239	
Qy	241 RNLEKFADEADIVRLYNQELVERWNLAASDILRISMLKEDGGVYLDVDILPGIQDILFKS 300	
Db	240 RANSILFTEQELLNYSQELLNRGNLAASDIVRLALKNFGVYLDVDMPLGHSIDLFKT 299	
Qy	301 INKPDSTINTSWEMIKLEAINMKYKEYIPGYTSKRFMDLDEBVEQSFESALSSKSDKSEIF 360	
Db	300 ISRPSSIGLDREWEMIKLEAINMKYKYNINYSSENFDKLDQOLKDNFKLIIESKSEKSEIF 359	
Qy	361 LPDDIKVSPLEVIKAPANNVINQALISLSDKSCDILVINOIKRYKILNDNLNPSINE 420	

Db	360 SKLENLVSDLEIKIAPALGSVINQALISKQGSYLTNLVIEQVKNRYQFLNHLNPAIES 419	
Qy	421 GTDFTNTMKIPLSKLASISNEDNMFMKITNYLVKGFADPVESTINLSGPGVYTGAYQD 480	
Db	420 DNNFTDTTKIFHDSLNSAENSMTLKIAPLQVGFMEPEASTISLSGPGAYASAYD 479	
Qy	481 LLMEFKDSTNHLLEPELRNFERPEPKTKISQLTCEITSLWSFNQARAKSOFEEYKGYFE 540	
Db	480 FINQENTIEKTLKASDLIEFPENNLSQLTQCEINLSWFSFQASAKYQFEKVRDYG 539	
Qy	541 GALGEDNDLPQAQNTVLVDKDY-VSKTILSS--MKTRNKYIHYIVQLQGDKIISYASCNL 597	
Db	540 GSLSDNGVDENKNTALDKYLLNNKIPSNVVEAGSKYVHYIQLQGGDISVEATCNL 599	
Qy	598 FSKDPYSILYQKNIEGSTAYVYVAD--ARIKEIDKYRIPQVISNRKNIKUTFIHGK 655	
Db	600 FSKNPKNSIIQRNM--NESAKSYFLSDGSGESILELNKYRIPERLKNKEKVKVTFIHGK 657	
Qy	656 SEFNTDTFANLDVDSLSEIETILNLAADISPKYIEINLLGCMFYSISAEETYPGKL 715	
Db	658 DEFNTSEFARLSVDSLSEIISFLDTIKLDISPKNVEVNLGCMFYSIDFNVEETYPGKL 717	
Qy	716 LLKIKDRVSELMPSISODSITVSANOYEVRINERKREILDHSGKWINKEESIIKDISK 775	
Db	718 LLSIMDKITSTLPDVKNKNSITIGANOYEVRINSEKRELLAHSKWINKEEAINMSDLSK 777	
Qy	776 EYISFNPKENKIIIVKSKYLHELSTLQEIERNANSSDIDLEKKVMLTECEINVASNIDRQ 835	
Db	778 EYIFPDSIDNKLKAKSKNIPGLASISEDIKTLLDASVSDPTKFIILNKLNISSIGDY 837	
Qy	836 IVEGRIEAKNLTSDSINYINKNEPKLIESDSYDIKHQGLDSDHIFISPEDISKTEG 895	
Db	838 IYVEKLEPVKNIHNSIDDLIDFENLLENVSDLEYELKLNLDKYLISPEDISKNST 897	
Qy	896 FRIRPINKETGNSFIETEKEIFSEYATHISKEISNKTIFDNVNGKLVKKNLDAHE 955	
Db	898 YSVRFINKSGESVYVETEKEIFSKYSEHITKEISTKNSIITDVNGLLDNILQDHTSQ 957	
Qy	956 VNTLNSAFFIOSLIEYNTTKESLNSLVAMKVQVYQALFSTGLNTITDASKVVELVSTAL 1015	
Db	958 VNTLNAFFIOSLIDYSSNKVDLNDLSTSVKQVYQALFSTGLNTIYDSIQVLNLSNAV 1017	
Qy	1016 DETID 1020	
Db	1018 NDTIN 1022	
	RESULT 6	
ID	ABO14444	
XX	ABO14444 standard; protein; 3169 AA.	
XX	AC ABO14444;	
XX	DT 23-OCT-2003 (revised)	
DT	21-AUG-2003 (first entry)	
XX	E. coli 0157 clostridial cytotoxin-like protein o3169.	
DE	OZID; acute haemorrhagic colitis; haemolytic uraemic syndrome;	
KW	food poisoning; clostridial toxin-like protein.	
XX		
OS	Escherichia coli; strain 0157:H7.	
XX		
PN	US2003023075-A1.	
XX		
PD	30-JAN-2003.	
XX		
PF	01-APR-2002; 2002US-00114170.	
XX		
PR	04-DEC-1998; 98US-0110955P.	
PR	03-DEC-1999; 99US-00453702.	
XX		
PA	(BLAT/) BLATTNER F R.	

PA (BURL/) BURLAND V D.
PA (PERN/) PERNA N T.
PA (PLUN/) PLUNKETT G.
PA (WELC/) WELCH R.
XX
PI Blattner FR, Burland VD, Perna NT, Plunkett G, Welch R;
XX
XX WPI; 2003-479497/45.
DR N-PSDB; ACD19243.
XX
XX New DNA sequences from *Escherichia coli* strain O157:H7, useful for
PT detecting *E. coli* O157:H7 in a sample, or in designing diagnostic probes
PT which can be used to distinguish strain O157:H7 from strain K12 using
PT molecular techniques.
XX
XX Disclosure; SEQ ID NO 257; 33pp; English.
XX
XX The invention relates to an isolated DNA molecule comprising an *E. coli*
CC strain O157:H7 sequence selected from a clostridial cytotoxin-like gene,
CC a urease gene cluster, a RIX toxin-like gene cluster, a locus of
CC enterocyte effacement and 2 genes from its associated lymphocytic phage
CC 933W (a putative serine/threonine kinase and a tail fibre gene), *E. coli*
CC O157:H7 can cause food poisoning, specifically acute haemorrhagic colitis
CC (which can develop into haemolytic uraemic syndrome). Also included are
CC an isolated DNA molecule comprising a nucleotide sequence identical to at
CC least 25 contiguous nucleotides contained in DNA sequences selected from
CC ACD18988-ACD19242 (being 255 *E. coli* O157 DNA sequences which are not
CC found in *E. coli* K12), a recombinant DNA construction comprising the DNA
CC above and a method for detecting *E. coli* O157:H7 (ATCC 43895) in a sample
CC (or distinguishing between O157 and K12) using a probe derived from one
CC of the 255 sequences. The DNA sequences are useful in detecting *E. coli*
CC O157:H7 in a sample, for the early diagnosis of humans and livestock
CC infected with O157:H7, and in designing diagnostic probes which can be
CC used to distinguish strain O157:H7 from strain K12 using molecular
CC techniques. The present sequence is the clostridial toxin-like protein.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO at seqdata.uspto.gov/sequence.html?DocID=20030023075 (Updated on 23
CC -Oct-2003 to standardise OS field)
XX
XX Sequence 3169 AA;
XX
XX Query Match 8.1%; Score 422.5; DB 6; Length 3169;
XX Best Local Similarity 21.0%; Pred. No. 7.5e-14;
XX Matches 258; Conservative 197; Mismatches 450; Indels 323; Gaps 54;
XX
QY 14 YVKFRIT--QDEYVAINAL--EEYHNMSSESVVEYKYLKLDINNLDNLYNTYKSGR 68
DB 160 YIKIRTRGAEQDTTITQSLIINELLNGVDRNTI-PPQKISELNDIIHSYENMQIKNSR 218
QY 69 N--KALKPKFVILTMEVLELKNS-----LTPVEKNLHFTWIGQ 106
DB 219 KGTEILVKGELLSSLLINDNKGKQLSDNASKIINLLGIEYQSHKVDIEPTTHAVWAGA 278
QY 107 INDNTAINYNQKVDNSDYTKVFDNSAF-----LINTLKT----- 144
DB 279 PPNTFSYITAFNLTYKYDYTLWIDPNFAGAAKFSGLKNTAMNAYIMRURTNPHLAE 339
QY 145 -----IVESATNNTLE--SPRENLDPEFY-----NKFYRKMEIIVYDKQHF 187
DB 339 ENNEVLKIQNTQNEIEKFEKERLKELENYKSLTSEKFNVFLESIGMQDNYF 398
QY 188 DYYKSQIENPEFI-----IDNIIK--TYLSNEYSKDLALNKYIEESLUNKITANGN-- 238
DB 399 TYCISNGISNTDDISRLDFTNLVLSPEVQDFKSTVEKNKRD:DLKNTISQKFGDRF 458
QY 239 ---DINLEKFADEDLVRINQELVERWNLAAASDILRISMLKEDGGVYLDVILPGIQP 295
DB 459 QLRDINTLESFKKPQDYPFYQOEMLRWNYAAASQVRINILKEYGGIYDTDLIPAYSD 518
QY 296 DLFPKSNKPDISINTSWEMIKLEAIMKYK--EYIPG--YTSKNFMDLDEVQSFESALS 351
DB 519 KVSQIINE-KSDDKRRFEDLKURRISESILSIKGEKYSIKH-DGLDETTLNQLNNIL- 575

352 SKSDKSEIFLPDDDIKVSPLVKIAPANNVI-----NQALISIKDSYC 395
DB 576 SEIEK-----LATIDDY-FKPVETKVVDRDFPKFYQYQKWTENTWIRGNMNFMLTKGSKC 630
QY 396 SLVINOIKNRYKIILNDNLNPSINEGTDFNTMTKIFSDKLASINE-----DNM 444
DB 631 IDFIISGQKKYLEL-QRIRDNISVNNLFYTT-----EDLSLNNVAIGGIPAKKYLEHG 684
QY 445 MFMIKITNLYKVGFPADVRSTINLSGP----- 471
DB 685 LF-----SEYRQDGTIPYVVVSTLINISGPDWIMQMKYKSLGRIGEVHIKONKLSDVNLF 740
QY 472 GVTYQAYQDLMF-----KONSTNIHLEPE-----LRNPEPKTKISQLT-----EQEITS 518
DB 741 GYVASSNKDKSNFNLNPFVSGINDITPDESSWVRNNDINKILFEKINCHVPEKLPTS 800
QY 519 LMSFQARAKSOFEEYKGYREGALGEDNDLFAQNTVLDKDYVSKILSSMKTENKEVI 578
DB 801 LY-----YEIDRSRFFQGW-----DNKSIKHTVTEINKDLI--KQINLLTSSNDV 844
QY 579 HVIYVQLQGDKISYEASCNLFSDPYSSILYQKNIEGSETAYYYVADAEIKEDKYRIPY 638
DB 845 KLLIKL--DRELYA:SSKI-----DNPLARSIRTLOQLANYVTSTFEPENTINFIY 896
QY 639 QISNKN-----IKLTFIGHGKSEBFTDTPANLDVS-----LSSEIET 677
DB 897 DFYRKQDQLLSA:KL-----FSRNDADTKIIVVYNSVMEKXVFLREVISC 942
QY 678 ILMKADISPKYIEINLLGCMFYSISAEITYGKLLKTKDRVSELMPSISODSI-- 735
DB 943 VLRSKKVD--SYINEN-----KQNSKEDA--GAURDYAKLKKKELFSLMDDDDGYKK 990
QY 736 TVSANOVEVRIIEEGKREILDHSGKWINKEESI:KDISSKEYISFNP-----K 783
DB 991 IITNAY-----IKERDKL-----SGIYIENSLIISGHESFDIIRSNQHEWGLSTVEQFK 1042
QY 784 ENKIIKVSRYLHLSLTLQEIIRNANSSDIDLEKKVMLTECEINVAS-----NIDRQIVE 838
DB 1043 KPEFYKSE-LSSAKSIFDDIKNK-YITDPETKRNVLVHQLSDIKERIAFLDISHYAYP 1100
QY 839 GRIEEAKNLTSDSINYKNEFKLIESI-----SDSLYDLKHQNGL 878
DB 1101 GSLEKQLQS---GYVFSIDNIIAEYLLASYGVSHGVVYVAPSPDKLLELRHTK 1156
QY 879 DSHFISFEDISKTENGFRIRFINKETGSI---FLETEKEIFSEVATHISKEISNIKT 935
DB 1157 SNSEWI--EKITP-----YVVDILSDNVNVRPPLSEQOKILNDIKLEISKVS---EQ 1207
QY 936 IFDVMNGKLVKYNLDAAEVNTLNGAFFIQSLIEYNTTK-----ESLSNLSV----- 983
DB 1208 YFMKLTQKSSVIGIKYVDVDFRYENLFLSLP:INQNLILPFMYRYFEMLYDIHIGIEN 1267
QY 984 -AMKVQVYQAQLFSTGLNTITDASKVVEL 1010
DB 1268 KANREFIYKFSLLNDFLINDERVNL 1295
XX
XX RESULT 7
XX AAR07503
XX ID AAR07503 standard; protein; 1254 AA.
XX
XX AC AAR07503;
XX
XX XX 25-MAR-2003 (revised)
XX DT 06-FEB-1991 (first entry)
XX
XX DE Merozite apical-end-localised protein (MABP) insert 5.3.
XX KW Malaria; vaccine; Duffy blood group.
XX OS Plasmodium vivax.

Qy	609	QKNIEGS--ETAYVYVADAE-IKEIDKYRIPQVIGNKENIKLTFIGHGKSEFT-----660
Db	507	ITNIEGALKESGNYEIGLEKUEEIGKGRKLKVDITKKSINST-VGNFSSLFNNFELNQ 565
Qy	661	-DTPANLD-----VDSLSEIETILNLAKADI-----SPKYI-----FINLL 696
Db	566	YDFNKINDYENKMGIEYNEFEGLNKISENLRNASENTSDYNSAKTLRLEAQKEKVULL 625
Qy	697	G-----CNMFYSISABETVPGKLLLKIKORVSELMPSISQDSITVSANQYEVRAINBEGKR 752
Db	626	NKEEENAKYLRDVKVYESF--RFIPFNKESLDKINEMIKKEQLTVNEGHGNGVQLVENIK 683
Qy	753	EILDHSGKWINKESIIKOISSKEYISFNPKENKIIIVKSKYLHELSTLLQEIIRNNANSSD 812
Db	684	ELVDE-----NNLSDLKQATGKN-----EEIQKITHSTLKNKAKTILGHVDTSAKYVG 732
Qy	813	IDLEKKVMLTE---CEINVASNIDRQIVBGRTEEAKNLTSDGINVIKNEFKLIESISD- 867
Db	733	IKITPELALTELGDAKLKTQAELEKFSKNVVLTEENMSKNT-----NELDVHKNIQDA 787
Qy	868	---SLYDLKHQNGLDSDHFSFEDISK-TENG---FRIRFINKETGNSIFITEKEIFS 919
Db	788	YKVALEILAHSDRIDTKQ----KDSKSLIEMGNOIVLKVVLINQYKKNKISIIKSKEEAVS 843
Qy	920	EYATHISK---EISNI---KDTIFDNVNGKLVKVKVNLDAAEVNTLNSAFFIQ-----966
Db	844	VKIGNVSKHSELSKITCSKSYDNI-----IALEKQTELQNLRSFTQEKTNNTNSDS 896
Qy	967	SLLEYNTTKESLNSLVAMK-----VQVYAQLFSTGLNTITTDASKVVVELV 1011
Db	897	KLEKIKTDFESLKNALKTLEGEVNALKASSDNHEHVQSKSEPNPALSEIEKEETDIDSL 956
Qy	1012	STALDETI 1019
Db	957	NTALDELL 964
RESULT 8		
AAW24575		
ID	AAW24575	standard; protein; 1254 AA.
XX	AC	AAW24575;
XX	AC	
DT	25-MAR-2003	(revised)
DT	10-NOV-1997	(first entry)
XX	DE	Merozoite apical-end protein clone 5.3.
XX	KW	Merozoite apical-end protein; MAEP; Plasmodium vivax; antigen; malaria;
XX	KW	human; erythrocyte; antimalarial vaccine; antibody; blood stage parasite;
XX	KW	Duffy blood group antigen; red blood cell; therapy.
XX	OS	Plasmodium vivax.
XX	PN	US5646247-A.
XX	PN	
PD	08-JUL-1997.	
XX	PF	04-OCT-1991; 91US-00792865.
XX	PR	05-APR-1989; 89US-00334041.
PR	06-APR-1989;	89US-00334270.
PR	03-APR-1990;	90MO-US001849.
PR	02-NOV-1990;	90US-00608639.
XX	FA	(UTNY) UNIV NEW YORK STATE.
XX	PI	Galinski MR, Barnwell JW;
XX	DR	WPI; 1997-362995/33.
DR	N-PSDB;	AAT80072.
XX	XX	Plasmodium merozoite apical end protein - useful as antigen for

PT	production of anti-malarial vaccines.	
XX	Claim 1; Col 29-38; 68pp; English.	
PS	AAW24575 and AAW24576 represent the merozoite apical end proteins (MAEP)	609 QXNIEGS--ETAYYYVADAE-IKEIDKYRIPYOISNKRNIKLTFIGHGKSEFT----- 660
XX	isolated from two different Plasmodium vivax strains. These proteins are	609 QXNIEGS--ETAYYYVADAE-IKEIDKYRIPYOISNKRNIKLTFIGHGKSEFT----- 660
CC	the antigens of the invention, and immunoreact with antibodies against a	507 ITNIEGALKESGNGVEIGFLEKLEIGKVRKLVDTITKKSINST-VGNFSLFNFDLQ 565
CC	native MAEP sequence. P. vivax is one of the four malarial species that	507 ITNIEGALKESGNGVEIGFLEKLEIGKVRKLVDTITKKSINST-VGNFSLFNFDLQ 565
CC	infects humans, and is difficult to target for a vaccine, as it cannot be	661 -DTFANLD-----VDSLSEIETILNLAADI-----SPKYL-----EINLL 696
CC	cultured in vitro. The preinvasion orientation of malarial merozoites	661 -DTFANLD-----VDSLSEIETILNLAADI-----SPKYL-----EINLL 696
CC	indicates that the apical end plays an important role in the invasion	566 YDFNKNINDYENMGGEIYNEFEGSLNKISENLNASENTSYNSAKTLRLAQKKNLL 625
CC	process. The MAEP protein binds to the surface of susceptible	566 YDFNKNINDYENMGGEIYNEFEGSLNKISENLNASENTSYNSAKTLRLAQKKNLL 625
CC	erythrocytes from P. vivax susceptible humans and primates, and also	697 G----CNMFYSISAEETYPGKLLKIKDRVSELMPSISQDSITVSANQYEVRIEENK 752
CC	binds to rabbit erythrocytes. The antigen can be used for the production	697 G----CNMFYSISAEETYPGKLLKIKDRVSELMPSISQDSITVSANQYEVRIEENK 752
CC	of antimalarial vaccines. The antigens are involved in the invasion	626 NKEEBANKYLRDVKKVESF--RFIFNMKESLDKINEMIKKQLTVEGHGNNVQLVENIK 683
CC	process, and are immunohistochemically reactive with antibodies raised against	626 NKEEBANKYLRDVKKVESF--RFIFNMKESLDKINEMIKKQLTVEGHGNNVQLVENIK 683
CC	malaria (particularly P. vivax) blood stage parasites. Synthetic	753 EILDHSGKWKNEEIIKIDISKEYISNPENKLIKVKLYLHELSTILLOIRNANSSD 812
CC	proteins, polypeptides, peptide fragments and analogues of these antigens	753 EILDHSGKWKNEEIIKIDISKEYISNPENKLIKVKLYLHELSTILLOIRNANSSD 812
CC	can be used similarly. As the antigens specifically bind to a Duffy blood	684 ELVDE-----NNLSILKQATGKN-----BEIQKITHSTILKNKAKTLGHVDTSAKTVG 732
CC	group antigen (the antigen present on the surface of susceptible	684 ELVDE-----NNLSILKQATGKN-----BEIQKITHSTILKNKAKTLGHVDTSAKTVG 732
CC	mammalian red blood cells), and are necessary in the process of invasion	813 IDLEKKWLTE-----CEINVASNIDRQIVEGRIBKXLTSDSINYIKNEFKLIESID- 867
CC	of red blood cells by merozoites, they can be used to inhibit the	813 IDLEKKWLTE-----CEINVASNIDRQIVEGRIBKXLTSDSINYIKNEFKLIESID- 867
CC	invasion of red blood cells by a malarial organism. The antigens can also	733 IKITPELALTELLGDAKLTAQELFESKNNVLENNKNT-----NELDVHKNIQDA 787
CC	be used in a method for inhibiting invasion of susceptible mammalian	733 IKITPELALTELLGDAKLTAQELFESKNNVLENNKNT-----NELDVHKNIQDA 787
CC	blood cells by malarial merozoites, and in a method for inhibiting the	868 ---SLYDLKHQGLDSDSHFISFEDISK-TENG-----PRIFINKETGNSIETETEKEIFS 919
CC	propagation of a malarial organism in susceptible red blood cells.	868 ---SLYDLKHQGLDSDSHFISFEDISK-TENG-----PRIFINKETGNSIETETEKEIFS 919
CC	(Updated on 25-MAR-2003 to correct PF field.)	788 YKVALEILAHSDIETKQ-----KDSKSLIEMGNQIYLVKVLINQYKNKISSIKSEEAVS 843
XX	Sequence 1254 AA;	920 EYATHISK---EISNI---KDTIFDNNVGNKLVKKNVNLDAHEVNTLNLSAFFIQ----- 966
QY	Query Match 5.7%; Score 294; DB 2; Length 1254;	844 VKIGNVSKHSELSKITGSDKSYDNI-----IALEKQTELQNLNRSFTOEKTNNSDS 896
Best Local Similarity 19.6%; Pred. No. 2.1e-07;		967 SLIEVNTTKESLSNLSVAMK-----VOVYAOQLFSTGLNTITDASKVVELV 1011
Matches 213; Conservative 198; Mismatches 408; Indels 269; Gaps 48;		967 SLIEVNTTKESLSNLSVAMK-----VOVYAOQLFSTGLNTITDASKVVELV 1011
64 KSGRNKALKPKFVLTVEVLELKNLSLTPVEKNLHFHWIGQINDTAINVQWQDVNS 123		897 KLEKINTDFESLKNALKTLEGEVNLKASSNHEHVQSKSEVPNPALSEIKEKETDIDSL 956
14 KKSIEKAYEKMGN--TLKELEKMD--EKN-----IEKEVEEAQIQY----- 52		1012 STALDETI 1019
124 DTVYKVFVDSNAFLNTLTKKIVTESATNTNLTESFRENLDPEFDYKFKRMELIYDKQ 183		957 NTADELLE 964
53 ---KRFIDHVV-----NLMNDEVEKSKIVMEKIB-LYKKE 84		
184 KHFIDYKSOEENPEFIIDNIITKYLGNESKOLEALNKVIEESLN-----KITANNGN 238		
85 ---IDELIKQNEYKQGTSPF---YYTEQNSATQSKAK-IEQFINIATTKGTSTSQ 137		
239 DIRNLEKFADE-----DLVRLYNQELVE-RWNLAASDILRISMLKEDGGVYLDVILPG 292		
138 DINELESIKEEVHKQLQVQBSNWSWEMRKQILSMKOLLILN----- 180		
293 IQPDLFKSINKPDSITNTSWEMIKLEAMKYKEYIPGVYTKNFDMLDEEVORSEFESALSS 352		
181 -----NSETIAKEISNNTQNALGFRENAKTK-----LNKTDLELQORVAAMIEEA 224		
353 KSDKSEIFLPDDIKVSPLEVKVIAFANNSVINQA-----LISLKD--SYGSDLVINQIK 404		
225 KAHKNNDIALEDAQIDTEVSKIEQINREIMNKDEIKSYLSEIKYKDKCTTISNSKR 284		
405 NRYKI-----LNDNLNPSINEGTDF-----NTTWKIFPSOKLA 436		
285 GDKLEFLEKFKPNEBESNKNVINEINENRSEQYLKDIEDAKQASTKVELFKHET 344		
437 SISR--EDNMFMPIKITNYLKVGFADPVRSTINLSGPGVYT--GAYQDLMLFKONSTNIH 492		
345 TISNIFKSEILGVETKSKINKAEDIMKEIERHNSBIQTQVKGQENLKNLNEPHNYD 404		
493 LLEPELRNEFFKTKISQITQEIITSLMSFNQARAKSQFEEYKGYFFGALGE---DDNL 549		
405 NAEDELNDKSTNAKY--LIETNLESV-----KXNLSEITNIKQ-----GEKITYSKAK 451		
550 DPAQNTVLQDVYSKILSKMTRNKEYIHYIVQLGGDKISYEASCNLPSKDPVSSI-LY 608		
452 DIMQKIKATSENTAETLEKVKDDQSNVNYL-----NOITERNLIVTERKRLNGIDST 506		

RESULT 9	
ABU48941	
ID ABU48941 standard; protein; 4688 AA.	
XX	ABU48941;
XX	AC
XX	DT 19-JUN-2003 (first entry)
XX	DE Protein encoded by Prokaryotic essential gene #34468.
XX	XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX	OS Ureaplasma urealyticum.
XX	XX WO200277183-A2.
XX	XX 03-OCT-2002.
XX	XX 21-MAR-2002; 2002WO-US009107.
XX	XX 21-MAR-2001; 2001US-00815242.
XX	XX 06-SEP-2001; 2001US-00948993.
XX	XX 28-OCT-2001; 2001US-0342923P.
XX	XX 08-FEB-2002; 2002US-00072851.
XX	XX 06-MAR-2002; 2002US-0362699P.
XX	XX (ELIT-) ELITRA PHARM INC.
XX	XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX	XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX	XX WPI; 2003-029926/02.
XX	XX N-PSDE; ACA52811.
XX	XX New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 76865; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: the sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX
 SQ Sequence 4688 AA;
 Query Match 5.6%; Score 290.5; DB 6; Length 4688;
 Best Local Similarity 18.8%; Pred. No. 1.8e-06;
 Matches 252; Conservative 220; Mismatches 433; Indels 437; Gaps 65;

QY 13 VYKFRIOEDYVAINLAJE-EYHNMSESVVEKYLKIDINLTDNYLNT---YKSGR 68
 Db 403 LYEVIKIKTNEVDVNLKQIPYHRS-----INNLSNALNTPYQYTKGD 449
 QY 69 NKALKKFEKLTWE-----VLELKNSS-----LTPVEKNLHFIWIGGOINDTAINYNQKD 120
 Db 450 INLIKAVPYVYVQVYGFQDQNOEHQILAKVKKDGTAFDTGALNNNSYSLDKIVS 509
 QY 121 VNSDYTVKVFYDGNAFINTLTKTIVESATNTLESFR-----ENL----- 161
 Db 510 VSNPQNVLV---SNFDLTSKQKQLIKKPAANASVDTSTKTOILENLNLDINQKLVATFV 566
 QY 162 --NDPEPDYKFKRMEIYDKQ---KHFDYKQSIEN----- 197
 Db 567 DNDKKEYKVAVNDQNNKIIFSDNLPKGYI-YHLAKVENNDLNKVINLNDPDLKDIID 625
 QY 198 -----PEFIID-----IKTVLSNEYSKDL---EALNKY-----IEESLN 230
 Db 626 KRDLNLIDSHPDFDYNDGNLEIHTQLANDLNDLQKALNNANVKGIVVDQDGIHEID 685
 QY 231 KITANNGNDIRNLEKFADELVR--LYN-QELVERWNLAASDILRISMLKEDGGVYLDV 287
 Db 686 VSDANGKVIPTKGLANNDPTKPIYITLKKVYKQNNQPNIDLISEEQSGDNHISFK- 744
 QY 288 DILPGIOPDLFKSINKPD-----SITNTSMEWIKLEALMKYKEYIPGYTSKNFDMLEDEVO 343
 Db 745 -----KPTTAKTKENDDYEISFNSPLANKIKLTKFTDD-----NNTWTKTVEAS 791
 QY 344 RSPESALSKSKSEIFLPP-----LDDIKV-----SPL----- 371

Db 792 IGLDGKAIKFTSDDAIFAPDHPKVTLTKEADNKKVANIDEISPLDRIVNKKQNGNVNADN 851
 QY 372 --EVKIAFANNYSI-----NOALISLKDSYCSDLVINQIKRY-----KI--- 409
 Db 852 KHEFKIPDQKNDLTAVYKDKKNNETHVPTKDDKGKVIWPNNNLFDPNKIYDFDKVD 911
 QY 410 LNDNLNPSINEGDFNTWKIFSD-----KLASISN-EDNMFMKIKITNVLKVG 457
 Db 912 LNEYPNKTILDRNSINKDVSAINDGVQDQARKLVKAPAVSNVTVAINFQVNLFDMLKUS 971
 QY 458 ----RAPDVRSTINLSGPGVYTGAYQ-----DLMFKDN 487
 Db 972 YNQFALTUTAKVNDLNDTOKYIATYDPKTNVYKLNDFTHLDANTKYVVDVLELNNK 1031
 QY 488 STNHLLEPELRNFEPPKTKISOLTEQEITSLSMFSNQAARAKSQ-----FEYKKGYPFG 541
 Db 1032 EKPIKLIKDVLPFEFTTSATTIN---PPIWTKEDVVVKTNNNDITTFEIDDK---DN 1084
 QY 542 ALGEDDNLDPQANTVLDKDYVSKILLSSMKTRKNEYIHIVQOGDKISYEASCNLFPSKD 601
 Db 1085 ILKNDOKI-YAQLALMDDDLDDTDVINPLVFSKT-----NKIASINGLDLKGNS 1133
 QY 602 PYSILYKQKIESETAYYYYVADAEIK-----EIDKYRIPYQISNKRNIKLTFIGHGK 655
 Db 1134 KYST-----KNL-----YYLNDQNKQVYLFKXNDVTKYB-QHFTNPKINLJSP---NK 1177
 QY 656 SEFWTDTF---ANL-----DVDLSSEIETILNLAKADISPKY 690
 Db 1178 SAVEQDIFADHANLFDYKDYDOKLRINEVDKVIYQINIDNKLQEL-IGYKV-VANNK 1234
 QY 691 IEINLLGNNFYSISABETYPGKLLKIKDRVSELMPS-----ISQOSITV 737
 Db 1235 IKFNLVG-----LKEKTIYVIKKLEALNKSASSIVNSEFLLDPTNFTSNKNTLV 1287
 QY 738 SANCY-----EVRINEEGKREILDHSGKWI-NKEBSIILKDISKEY 777
 Db 1288 GLNSIDNNNDQTFPIINAKINIGDDFQDNOQVKLIYVSNNDNKEIKSAVTLIKQRNYQF 1347
 QY 778 ISFPNPKNKLIVKSKY-----LHELSTLLQEI-----ENNA-----NSSDIDLEKVM 821
 Db 1348 EFSNLKLNRLYTFKSIYVETNNQTLHKLDTLTHQFSINPNNNAVSLKKNNTNIEITKRLV 1407
 QY 822 TECEINVASINDROIVEGRIE--EAKNL--TSDSINYIKNEFKLIESISDLYDLKHQNG 877
 Db 1408 -----NNDQSLISAKIEVDIDVNLTAQPNIVYO-----LENSDANNKLATN 1454
 QY 878 L-----DSSHFIISPEDI-----SKTENGFRIRPINKETONSIF----- 910
 Db 1455 AQVVVENNKKFLKFDLVLKINQNVYIKESIFSNKPTNAY-FNFTNNKTNIVSYDEQN 1513
 QY 911 -----IETEKEIF-SEYATHISKEISNFK-----DT 935
 Db 1514 KISLSNNINPTSYKPTQNTINKDKSVNIDVDLQVDKQLLANQYLRKLKQLNDNKTVMWTD 1573
 QY 936 IFDVGNGKLVKKNLDAAEVNTLNSAFFQSLI---EYNTTKESLSN-LSVAMKVQVYA 991
 Db 1574 ILFNNAKISPKLS-NLIH-----NRAYLEGLYFDQNSVNDMTNNTQISFNSKIHKPK 1627
 QY 992 QLFSTGLNTIT-DASKVVVELVS 1012
 Db 1628 IEFPSLTTINDTNNAIKTVS 1649
 RESULT 10
 AAB18171
 ID AAB18171 standard; protein; 1979 AA.
 XX
 AC AAB18171;
 XX
 DT 07-NOV-2000 (first entry)
 XX
 DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:28.
 XX

AAB18195
 ID AAB18195 standard; protein; 1516 AA.
 XX
 AC AAB18195;
 XX
 DT 07-NOV-2000 (first entry)
 XX
 DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:52.
 XX
 KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 XX
 KW antimalarial; malaria; protozoacide; infection; insecticide.
 XX
 OS Plasmodium falciparum.
 XX
 PN WO200025728-A2.
 XX
 XX 11-MAY-2000.
 XX
 XX 05-NOV-1999; 99WO-US026796.
 XX
 XX 05-NOV-1998; 98US-0107131P.
 XX
 XX (HOFF// HOFFMAN S.
 PA (CARU// CARUCCI D.
 PA (GARD// GARDNER M.
 PA (VENT// VENTER J C.
 XX
 XX Hoffman S, Carucci D, Gardner M, Venter JC;
 PI
 XX
 XX WPI; 2000-365347/31.
 XX
 XX
 PT Proteins encoded by chromosome 2 of the human malarial parasite,
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the
 PT diagnosis of P.falciparum infection.
 XX
 XX Disclosure; Page 120-124; 577pp; English.
 XX
 CC The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
 CC vaccines against P. falciparum infection comprising (I) or (II). (I) and
 CC (II) are useful for the development of vaccines against P. falciparum
 CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to
 CC immunogens comprising the sequences of (I), are useful in the detection
 CC of infection with P. falciparum. Furthermore, (I) (especially when they
 CC are rifins or secreted or membrane proteins) can aid the identification
 CC of drugs to treat or prevent P. falciparum infection, or they can be used
 CC to identify drug resistance in P. falciparum. Sequencing of the
 CC Plasmodium chromosome 2 and the subsequent identification of proteins
 CC encoded by it will help to expand our understanding of parasite biology,
 CC a process hampered by the complexity of the parasitic lifecycle, and
 CC provide new targets for vaccine and drug development. Parasite resistance
 CC to drugs and mosquito resistance to insecticides have led to a resurgence
 CC of malaria in many parts of the world, and there is a pressing need for
 CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
 CC represent nucleotide and protein sequences given in the present
 CC invention, but which are not specifically mentioned within the
 CC specification
 XX
 SQ Sequence 1516 AA;
 Query Match 5.5%; Score 283; DB 3; Length 1516;
 Best Local Similarity 19.2%; Pred. No. 1.1e-06;
 Matches 230; Conservative 187; Mismatches 351; Indels 432; Gaps 57;
 6 KAOLQKMYVFRQDEYVAILNALBEYHNMSSSVYVYKLLKQINNLT-----NYL 60
 131 KTAELVYVYQ-----RLLNSMKSNNKNNNS-----NNEDKTNPKT 173
 61 NTKYKSG-----RNKALKPKPEYL-----TMEVLE 85
 174 NTQNSNTQNSNTPNKINADISKSSLIQYDDIKEDKSLNSLVEHGVNVPVSKDVL 233

RESULT 12	Query Match	5.4%	Score 279;	DB 6;	Length 1178;
ABU24510	Best Local Similarity	21.0%	Pred. No. 1.3e-06;		
ID ABU24510 standard; protein; 1178 AA.	Matches 237; Conservative	193;	Mismatches 385;	Indels 316;	Gaps 58;
XX AC	ABU24510;				
XX AC					
XX DT	19-JUN-2003 (first entry)				
XX XX	Protein encoded by Prokaryotic essential gene #10037.				
DE XX	Antisense; prokaryotic essential gene; cell proliferation; drug design.				
XX KW	Clostridium botulinum.				
OS XX	WO200277183-A2.				
XX FN	03-OCT-2002.				
XX PD					
XX XX	21-MAR-2002; 2002WO-US009107.				
XX XX	21-MAR-2001; 2001US-00815242.				
PR PR	06-SEP-2001; 2001US-00948993.				
PR PR	25-OCT-2001; 2001US-0342923P.				
PR PR	08-FEB-2002; 2002US-00072851.				
PR PR	06-MAR-2002; 2002US-0362699P.				
XX XX	(ELIT-) ELITRA PHARM INC.				
XX XX					
XX PI	Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;				
PI PI	Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;				
XX XX					
DR DR	WPI; 2003-029926/02.				
XX XX	N-PSDB; ACA28380.				
PT PT	New antisense nucleic acids, useful for identifying proteins or screening				
PT PT	for homologous nucleic acids required for cellular proliferation to				
PT PT	isolate candidate molecules for rational drug discovery programs.				
XX XX					
XX PS	Claim 25; SEQ ID NO 52434; 1766pp; English.				
XX CC	The invention relates to an isolated nucleic acid comprising any one of				
CC CC	the 6213 antisense sequences given in the specification where expression				
CC CC	of the nucleic acid inhibits proliferation of a cell. Also included are:				
CC CC	(1) a vector comprising a promoter operably linked to the nucleic acid				
CC CC	encoding a polypeptide whose expression is inhibited by the antisense				
CC CC	nucleic acid; (2) a host cell containing the vector; (3) an isolated				
CC CC	polypeptide or its fragment whose expression is inhibited by the				
CC CC	antisense nucleic acid; (4) an antibody capable of specifically binding				
CC CC	the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular				
CC CC	proliferation or the activity of a gene in an operon required for				
CC CC	proliferation; (7) identifying a compound that influences the activity of				
CC CC	the gene product or that has an activity against a biological pathway				
CC CC	required for proliferation, or that inhibits cellular proliferation; (8)				
CC CC	identifying a gene required for cellular proliferation or the biological				
CC CC	pathway in which a proliferation-required gene or its gene product lies				
CC CC	or a gene on which the test compound that inhibits proliferation of an				
CC CC	organism acts; (9) manufacturing an antibiotic; (10) profiling a				
CC CC	compound's activity; (11) a culture comprising strains in which the gene				
CC CC	product is overexpressed or underexpressed; (12) determining the extent				
CC CC	to which each of the strains is present in a culture or collection of				
CC CC	strains; or (13) identifying the target of a compound that inhibits the				
CC CC	proliferation of an organism. The antisense nucleic acids are useful for				
CC CC	identifying proteins or screening for homologous nucleic acids required				
CC CC	for cellular proliferation to isolate candidate molecules for rational				
CC CC	drug discovery programs, or for screening homologous nucleic acids				
CC CC	required for proliferation in cells other than <i>S. aureus</i> , <i>S. typhimurium</i> ,				
CC CC	<i>K. pneumoniae</i> or <i>P. aeruginosa</i> . The present sequence is encoded by one of				
CC CC	the target prokaryotic essential genes. Note: The sequence data for this				
CC CC	patent did not form part of the printed specification, but was obtained				
CC CC	in electronic format directly from WIPO at				
XX XX	ftp.wipo.int/pub/published_pct_sequences				
XX SQ	Sequence 1178 AA;				

QY 918 -----FSEYATHISKEISNIKOTIFDNVNGKLVKVNLDAAHEVNTLNSAFFIOSLIEYNT 973
 Db 1046 KDITGRVALELD---CNGNTWRDDDFNGGIRPTNLSGGET-----FLTSLA----- 1091
 QY 974 TKESLSNLSVAMKQVQYQA-----LFTGLNTITDASKWELSTALD 1016
 Db 1092 -----LALALSSQILKGSPLFFFLDEGFGTLD--SELLDTVMTSLE 1133

RESULT 13

ID ABU25521
 ABU25521 standard; protein; 1182 AA.

AC ABU25521;
 DT 19-JUN-2003 (first entry)

DE Protein encoded by Prokaryotic essential gene #11048.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX Clostridium difficile.

OS WO200277183-A2.
 PN 03-OCT-2002.

PD 21-MAR-2002; 2002WO-US009107.
 PP 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.
 PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Walli D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

PI WPI; 2003-029526/02.
 DR N-PSDB; ACA29391.

XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX Claim 25; SEQ ID NO 53445; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation. (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1182 AA;

Query Match 5.3%; Score 276.5; DB 6; Length 1182;
 Best Local Similarity 20.7%; Pred. No. 1.7e-06;
 Matches 222; Conservative 205; Mismatches 372; Indels 275; Gaps 55;

QY 8 QLOQWVYKFRIOEDYVAILNALBEYHNHSSVVEKYLKLDINNLNDYNTYKKG 67
 Db 199 QLOQLKOLKSKL--SQYDGITE--EYNNINQELIELKQLE-KDKNALDLAQKSYEES- 252
 QY 68 RNKALKKFKVLTWMEVLELKNNSL---TPVENLHFIWIGQINDTAINYNQWQDVNS 123
 Db 253 --KTV--YESQLKLEKQELKRNELDLKNSIEKIALV-----ENDSNKKVDPY--IS 301
 QY 124 DYTQVYFDSNAFLINTLTKTIV-----ESATNNTLESFRENLD--PFDYNNKFKRGM 176
 Db 302 VONLEKKIDEDSPAVSNLEKKLAILNQLELVTKRYEKINKIKNEVPKLSSEKI-RLQQ 360
 QY 177 BIIVDKOKHFDYKSOIENPEFIDNIIKTLSNEYSKOL-----EALNKVIESLNK 231
 Db 361 AKLEBELVLLDLRELKELKESG---INLNTKVELEKVKQVSESERKDAVTKSKEVEGK 416
 QY 232 ITANNGNDIRNLKFADEDLVLYNQELVER--WNLAAASDILRLSMKEDGGVYLDVDIL 290
 Db 417 IDKVNISAEUKQKIFLAYEYKDYNNVLEBNOKLNLBEI-----LKTENINLKV--- 468
 QY 291 PGIQPDFKSIKPDSTINTSWEMIKLEAIMKYKEYIPGYTSKNFOMLDEEVQSPESAL 350
 Db 469 -----RYIDKNDVNWENLENLSLHLDVLKK-CPG--KSAULL---LKSEVTEL 513
 QY 351 SSKSDKSEIFLPLDDIKVSPLEKVIAPANNVINOALISKQSYCSLDLVINQIKNYKIL 410
 Db 514 RNK-----ANNTRENE-----IKKS-----SIQDELKII 537
 QY 411 NDNLNPSINEGTDFNT--TWKIFSDKLA-SISNEDNMFMKITNYLVKVPADPVRSTIN 467
 Db 538 LE-----SKPNTREVNLDNKLNNRKNRDLEKELEELKYLNL--ASELRREUK 586
 QY 468 LSGFGVYTG-----YQDLMFKNSTNIHLLEPRLNFFPKTKISQLTEQ---EITSLW 520
 Db 587 ENMPCPVCGSKHNDITKYDE--NISFVKEKLEKEKISIRNTEELNAKVSQYL 643
 QY 521 SFNOARAKSQPEYKKGYPFEGALGEDDNLDFQNTVLDKDYVSKTILSMKTRNKYIHY 580
 Db 644 SIERKTK-BLEDVK-----GKLGE-----IPSSQLLKLDEBQRKALLKSNIOE- 688
 QY 581 IVQLQGGKISYEASCNLFSKDPVSSILYQKNYBGSYAYVYVADAEIKEDKRYPIQI 640
 Db 689 -----WKEKESTENK--VTIAKEEKNIKEELKIRE 719
 QY 641 SNKENIKLTFIGHKSEFNTDTFANLDVPSLSSEIETILNLAKADISPKYIEINLGCNM 700
 Db 720 SLNNYKLT-----RDNLIEISLENKCK-KIKQYVLGLKTI---- 755
 QY 701 PSYISAEETYPQKLLLIK-----DRVSELMP-----ISQDSITVSANQYVRAINEE 749
 Db 756 -----TKVSNLLSKVKEIRENEKELELLSTGSHNLLKNRDSLDIKIRENESKLEH- 805
 QY 750 GKREILDHSGKWNKESIEIIOKDISSEYISFNPKENKIIVKSKYHELSTLQETRNMAN 809
 Db 806 --TEIELIKARELYVKKLSRDNKYKEVISITKGD-----LAKNLLHNVENICKILEQEE 859
 QY 810 SSDIDLEKKVMTCEINVASNIDROIVEGRIEAEAKNLTSD---SINYI--KNEFKLIES 864
 Db 860 SSKKLEBEQL--EYKKNIA---EKNIDGRUKTAQEQYKQKQSTINKULAENKESIYA 914

QY 865 ISDSLDP-----LKHONGLDSDHPSFISDISKTB--NGFRIRFINKTGNISFI 911
DB 915 VRALLDTPDAVKKLEIELEHE-----EQKULSFVAKSKSKLNG---RSIKKE-----YFE 964
QY 912 ETEKEIFSEYATHISKEISNIKDTIFDVGKLVKYNLDAHEVNTLNSAPFIQSLIEY 971
DB 965 QKDEIY-----NLKVEIGNISKDIGANQQLITLKQSLD---KINDFNKQL---KVVEH 1013
QY 972 NTKTESLNLVAMKVQVYALFSTG-----LMTITDASKVZVLVST 1013
DB 1014 NV--DLLEELDKVQGNKEFVYVSTNQLKXIALEBASKRLDGTIKGRVLEIDST 1065

RESULT 14
ABU24531
ID ABU24531 standard; protein; 1279 AA.
XX
AC ABU24531;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #10058.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Clostridium botulinum.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
XX
PR 06-SEP-2001; 2001US-00948993.
XX
PR 25-OCT-2001; 2001US-0342923P.
XX
PR 08-FEB-2002; 2002US-00072851.
XX
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI: 2003-029926/02.
XX
XX N-PSDB; AC28401.
XX
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 52455; 1765pp; English.

XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX

QY 25 VAINALBEYHN-MGESSVVEKYLKDKDINNLTNYLNTYKSGRNKALKKPK--EYJTM 81
DB 310 VELKNILDSYENEFIKKEFIFLYPSMKALSNLVILDKKYEAKKERDLIDFNDIEHLCL 369
QY 82 EYLEKNNSLTPEKNLHFIWIGGOI--NDTAINTINOWKOVNSDYTVKVFYDSNAFLIN 139
DB 370 SILTDKNSE-----GHIIPSDIALDYRKFAEV-----LIDEYQDSN--LVQ 409
QY 140 TLKKTIVE-----SATNNTLESFRE--NLNDPEFY--NKF-----YKRMET 178
DB 410 EVTMSVSRVKGYSFYNGQL-MFNEBEINLEPQICLDIPRMVGVGVKQSIYFRQA- 467
QY 179 IYDKOKHFIDYKSOIE-----NPEFIIDNIITKYSNE--YSKOLEALN 222
DB 468 ---KPEIFLDKNEYSEEGTKRKVKLPKPKRSKEVINGV--NYLPKQIMSKITIGELD 522
QY 223 KYIEESLNKITANNNDIRN-----LEKPAD--EDLVRLYN--QELVERNL---AA 267
DB 523 -YTEEEALVGSYGEEVKGGEIELCLMDKYEISEEVLKEYNVDEEALDNIQLEGLV 581
QY 268 ASDILRISMLKEDGV-----YLDVILFGIQDLPKSNKPKSIITNWSWE 313
DB 582 AKKIQLGVNLGGLGVFKLGEYRNLYQYRDIVL-----MRATSNWA 626
QY 314 MIKLEAIMKYKEVTPGYTSKNFMDLDEEVQSFESALSSKSKSEIFLPLDDIK-VSPLE 372
DB 627 PIFVEELA--KEGIPVFADTNGYFDTAIBKTIWSLLOI-----IDNPLQDIPLLSVLR 678
QY 373 VKIA-FANNSVINQALISLSDY--CSDLVINQIKNR-----YKILNDN 413
DB 679 SPIASTDDELDIRMVKNKITFYECWEIIRLYKNEKLDYVSFYIEDENKINKIKDM 738
QY 414 LNPSINEGTDFTNTMKIFSDKLASISNEDNMMPMKITNYLVKGVFPADYVRSINLSGPGV 473
DB 739 NEKLKNIKCSFIEKLKWEKSIHIDIDFIWFL-----YVETGYYG----- 780
QY 474 YTGAYQDLMPKDNSTNIHLLPELRFNFPFKTKISQLEBEITSLWSFNQARAKSQPEE 833
DB 781 YAGALQ--AGEORQANLRI-----FORAK--QYAKTSYKGLFNFINFKLKFS 827
QY 534 YKGYEGEGALGEDNDLDFQAQNTVLDKDYVSKKLSMKTNRKEYIHIVQLQDQKISYEA 593
DB 828 GDMGSAK--ILGENEV-----VRIMSHKSKGLEF--PVVILSG-----T 864
QY 594 SCNLFSDKDPYSSILYQKNEGSEYAYYVYVADAE-----IKEI-----DKY 634
DB 865 GKFNMTDLANKNLFHRDL-----GYGPDVDPERRIAVPSLVKNIKNKIRLETISEM 919
QY 635 RPYQTSNKNENIKLTPIGHCKSEFNFTDTANLVDVDSLSSEIETILNKAADISPKYIEIN 694
DB 920 RILYVALTRAREKLIITGL-----INNMDKTVDWNLNLS-----DKN 957
QY 695 LLGCNMFYSISAEETY---PGKLLKIKDRVSELMPSISQDSITVSANQYEVRIINEEGK 751

Query Match 5.3%; Score 275.5; DB 6; Length 1279;
Best Local Similarity 21.2%; Pred. NO. 2.2e-06;
Matches 239; Conservative 187; Mismatches 382; Indels 317; Gaps 62;
Sequence 1279 AA;

958 ----KVPEVAVMSGKTYLDWIGPALIKKHDAVS-----FREELKWTSP-L 997
752 REILDHSGKW----INKESEIIKDISKEY-ISPFPKMKI-IVKSKYLHELSTLLQIR 805
998 SNIVDDKSKWELWKNRLLKEKVEDEVEISEKIKETLWNLNLSNKEKIEKLSFKY 1057
806 NNANSSDI-----DLEKKVMTCEINAVASNIDR--QIVETR-----IEBAKNTSDSI 852
1058 KYDNASSIPTKLSVDVKQFIDKE-----NTBELFKLELRKPMFMEKKKISPSER 1112
853 NIKNEP--KLTESISDSYDLKHQ-NGLDSDHFSFEDISKTEGFRIR-RE----- 900
1113 GTIHLFMQHLDLKKAENEEDIKQINRLIEREFITYEQ-SKVINPYKILKFCRGELGKR 1171
901 -----INKETGSIIFETIKETIFSEYATHISKE-----ISNKTIDIFONVNGKLVKYNL 950
1172 ILNSNNVNKEMPPSIIPA-LEIYKELDKIYKDEKLIIOGVIDYCFEEDGLVLDYKT 1230
951 DAAHEVNTLNSAFFIOSLIEYNTKESLSNLSVAMKVQVYAGLFS 995
1231 DYVNDIEEKVRIEQ--IKY--YBEALNRIITGKNVVDKYLVLFS 1271

RESULT 15
AAB18294
ID AAB18294 standard; protein; 980 AA.
XX AAB18294;
XX 07-NOV-2000 (first entry)
XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:152.
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX antimalarial; malaria; protozoacide; infection; insecticide.
XX Plasmodium falciparum.
XX WO200025728-A2.
XX 11-MAY-2000.
XX 05-NOV-1999; 99WO-US026796.
XX 05-NOV-1998; 98US-0107131P.
XX (HOFF/) HOFFMAN S.
XX (CARU/) CARUCCI D.
XX (GARD/) GARDNER M.
XX (VENT/) VENTER J C.
XX Hoffman S, Carucci D, Gardner M, Venter JC;
XX WPI; 2000-365347/31.
XX Proteins encoded by chromosome 2 of the human malarial parasite,
XX Plasmodium falciparum, useful as antimalarial vaccines and in the
XX diagnosis of P.falciparum infection.
XX Disclosure; Page 362-365; 577pp; English.
XX The present invention describes proteins and their fragments (I) encoded
XX by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
XX Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
XX vaccines against P. falciparum infection comprising (I) or (II). (I) and
XX (II) are useful for the development of vaccines against P. falciparum
XX infection. (I) and polyclonal antisera or a monoclonal antibody raised to
XX immunogens comprising the sequences of (I), are useful in the detection
XX of infection with P. falciparum. Furthermore, (I) (especially when they
XX are fusions or secreted or membrane proteins) can aid the identification
XX of drugs to treat or prevent P. falciparum infection, or they can be used
XX to identify drug resistance in P. falciparum. Sequencing of the
XX Plasmodium chromosome 2 and the subsequent identification of proteins

CC encoded by it will help to expand our understanding of parasite biology,
CC a process hampered by the complexity of the parasitic lifecycle, and
CC provide new targets for vaccine and drug development. Parasite resistance
CC to drugs and mosquito resistance to insecticides have led to a resurgence
CC of malaria in many parts of the world, and there is a pressing need for
CC vaccines and new drugs. AAY70078 to AAY70287 and AAB18144 to AAB18352
CC represent nucleotide and protein sequences given in the present
CC invention, but which are not specifically mentioned within the
CC specification
XX
SQ Sequence 980 AA;

Query Match 5.2%; Score 272; DB 3; Length 980;
Best Local Similarity 20.4%; Pred. No. 2.4e-06;
Matches 224; Conservative 182; Mismatches 391; Indels 320; Gaps 52;

QY 6 KAOLQKVVY-VKPRIQDEYVAILNALAEYHNHNSSESVVEKYLKLDINNLTDNYLNTYK 64
DB 39 KEKGEKVIYELKELDNDE--KIINDLKKNYSY-----QVYKMDYERENNLINEN 89
QY 65 KSGENKALKPKFMYLTMEV---LELKNSLTTPV-EKMLHFIWIGQINDTAIYNQWKD 120
DB 90 K-----LKLFIENKMTVERGEMNKNKLEEMKQKNKELINNLINLIDISDELKNCIEQVNS 143
QY 121 VNSDYT-----VKVFYDSNAFLINTLTKTI---VES-----ATNNTLES 156
DB 144 VSRNMANVERKENIINELQILRMKN---DTWKRIKSFVEQEKVLKFKLYTINNDIFS 199
QY 157 FRENLDPEFDYKFKYKMEIIVYKOKHFID-----YVKSQIENPFIIDNIKT- 208
DB 200 KNEKLD-----MOKLNDVNEKYKIVVECLNNYKTEHKEQIEKKIERI--NTLKQN 249
QY 209 --YLSNEYSKDLALNKYIE-----ESLNKITANNNGDIRNEKFADED 250
DB 250 YYYIKVEYDLKKNKELEKNIHKGKLEHLSHCYEENQKLNKEEIKRRNSFIKNKDR--KID 307
QY 251 LVRLYNQELVERWNLAAASDILRISMLKEDGGVYLDVILPGIOPDLFKSINKPDSITNT 310
DB 308 LITNIENELLKK-----KEINNIKMEKQVKNNEQLLKDIDKENEKQNEHVNLQN- 360
QY 311 SWEMIKLEAIMKYKEYIPGYTSKNPMDLDEBQVSFESALSSKSKSEIFLDDIDKVP 370
DB 361 --ELIKRE-----LQNKCKSIDIEFCCKEKEDK-----IKNEDDL 394
QY 371 LEVKIAPANNVINOALISLSDSYCSDLVINQIKNRYKILNDNLNPSINDEGTDFNTMKI 430
DB 395 LEKKKCIEN--LKDELINIK-----KMKEDKMTNE-MDLNLSNKEVELNRINKT 441
QY 431 FSDKLASISNE-DNMFMKITYLKYGFADPVDVSTINLSGPGVY-TCAYO-----D 480
DB 442 YEKNIVELNNELDVIKKLNDDEBFLK---BEEKKNIDM---VYKIKEYEIQEKENE 494
QY 461 LLMFNDNSTNTHLEPELNRNPFPPKTKISQLTQETISLMSFNQARAKSQPEEYKGYFE 540
DB 495 IDSLAKKQNLHLVAKNE-----ELNEXEIIKKNYDK-EINMIIEQYNKKI-- 539
QY 541 GALGDDNLDPAQNTVLDKDYVSKILSSMKTRNKYTHYI-----VOLQDKISYEASC 595
DB 540 -----QEEKDMLNKKIKSMQDQTHKNQIEEMQFENKELKRLKNVCDMMQLQILKENEK 594
QY 596 NLFPSK-DPVSSITLYOKNIEGSETAYVYVYVADAIEKE-IDKYRIPYQISNK----- 643
DB 595 HMQEKVEEYKNNLKKQ-----DOELKNVIOEYDERIEIQKEMEDIVNDC 640
QY 644 -----RNILKTIHGHKSEFNFTDTPANLDVDSLSSEIETILNLAKADISPKYIENL 695
DB 641 EKLQAKANNKKLT-----TATNANNNNMLMDENLKEKDKKINDLAKMKEKKEEINK 694
QY 696 LG-----CNMPSYISAEETYP---GKLLKIKDRVSELNPSISQDSITVSA 739
DB 695 LVEEKSLESHVHKIQNEMSLIVEQNEKLKEEMGLSRIAID-MEEIKDMKEYEEXKK 753
QY 740 NQVEVRIINEEGREILDHSGKWINKEESIIDKISSKEVISFPNKENKIVKSKYLHELST 799

Db	754	NEERKQNEERKQNEERK--NEEKKQNEERK-----NEEKKKLEKXHQFE----	804
Qy	800	LLQEIERNANSDDIDLEKKVMTCEINVASNIDROIVEGRIBEAKNLTSDSINYIKNEF	859
Db	805	-----EEKERM--EYEHQKEDR---KKDKKKKGHSDEEKYNKKE	842
Qy	860	KLIESDSLSYDLKHQGLDDSHFISFEDISKTENGFRIRPINKETGNSIFIFIETEKEIFS	919
Db	843	KTKEKSSNILF-----DEEYIIQLEEL-----RDTGENCFI-----	873
Qy	920	EYATHISKEISNIKDTTFDNVNGKLVKKVNLDAHEVNTL-----	959
Db	874	-YLSLSKEL---DVIINKLKSDDALLN-DAFNKINLAITSWNIFNEENKEGDNITTV	927
Qy	960	-NSAFFIQSLIEYNTTKESLSNLSV--AMKVQVYQAQ	993
Db	928	ENTATEGNIITIDENTTEVENNEEVYKIFSVEKYDML	964

Search completed: April 1, 2004, 16:41:35
Job time : 50.75 secs

Db 1599 NINFDKFTLTKGKTLGVGFICDNNKNDIYFGEMKTSKSTIFSGNGRNVVVEPI 1658
Qy 1662 YHDDSGNISTVINSQKVLGIDRYVNVKVIAPNLYT 1700
Db 1659 YN-DPTGEDISTDFSYPLPYGIDYINKVLIAPDLYT 1696

RESULT 5
S55805
alpha-toxin - Clostridium novyi (ATCC 19402)
C:Species: Clostridium novyi
A:Variety: ATCC 19402
C:Date: 28-Oct-1996 #sequence revision 08-Nov-1996 #text change 05-Nov-1999
C:Accession: S55805; S71294; S71158; S4273; 140834; S4272
R:Hofmann, F.; Herrmann, A.; Habermann, E.; von Eichel-Streiber, C.
Mol. Gen. Genet. 247, 670-679, 1995
A:Title: Sequencing and analysis of the gene encoding the alpha-toxin of Clostridium novyi
A:Reference number: 140834; MUID:95342160; PMID:7616958
A:Accession: S55805
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-2178 <HOF>
A:Cross-references: EMBL:Z48636; NID:g728537; PIDN:CAA88565.1; PID:g755724
A:Accession: S71294
A:Molecule type: Protein
A:Residues: 1-15 <HOW>
R:Hofmann, F.
submitted to the EMBL Data Library, March 1995
A:Reference number: S71158
A:Accession: S71158
A:Molecule type: DNA
A:Residues: 1-1179, 'LKV', 1183, 'LVTHIGE', 1191-2178 <HOS>
A:Cross-references: EMBL:Z48636; NID:g728537; PIDN:CAA88565.1; PID:g755724
R:Hofmann, F.; Habermann, E.; von Eichel-Streiber, C.
submitted to the EMBL Data Library, July 1993
A:Description: Sequence analysis of Clostridium novyi alpha-toxin: a member of the family
A:Reference number: S44272
A:Accession: S44273
A:Molecule type: DNA
A:Residues: 1-243; 1204-2178 <HOA>
A:Cross-references: EMBL:Z23281
C:Genetics:
A:Gene: tcn-alpha
C:Superfamily: cpl repeat homology
C:Keywords: virulence factor
E:1880-1899/Domain: cpl repeat homology <CP12>

Query Match 24.4%; Score 2120.5; DB 2; Length 2178;
Best Local Similarity 31.3%; Pred. No. 1.1e-68;
Matches 552; Conservative 359; Mismatches 706; Indels 147; Gaps 46;

Qy 3 LVNKAQLOKVVYKFRIOEYVAILNALBEYH-NMSESSVVEKYLKXDINNLTNYLN 61
Db 2 LITREQLMKASIFPKKBEYNILDALENFNRIEGTSVKETIYKSLKSLNELVNYQT 61
Qy 62 TYKSGRNKALKKPEKYLTMVEVLKNNSLTP--VEKNLHFIMTGGQINDTAINYNOMK 119
Db 62 KYPSSGNLALFNEDSLSYSELRLIKNSRTSTIASKNLSFIMTGGPIQSQSELYNMWK 121
Qy 120 DVNSDYTVKPYDSNAFLINTLKTIVESAATNNLTESFRNLNDPDPYKFKRKRMBII 179
Db 122 MFNKDYNIRLPYDKNSLLVNTLKTATIQESSKVIIEQNSKILDTGTGHNKFPYSDRMKLI 181
Qy 180 YDKQKHFDYKSOIENPEFIIDNIITKYLINSYKSLDALNKYIESLNKITTANNND 239
Db 182 YRYKRELKMLYENMKQNN--VDDIINFILSNYFKYDIGNKQNNKNNKMIAGATD 238
Qy 240 IRNLEKFADEDLVRLYNOLVERWNLAAASDILIRISMLKEDGGYVLDVDPILGQPDLPK 299
Db 239 I-NTENILTNKLSYVYQELIQTNLAAASDILIRIALIKYGGYVCLDPLPGVNLSEFN 297
Qy 300 SINKPDSITNTSWEMIKLEAIMKYEYIPGYTSKNFDMLEDEVOQSFESALSSKSKSEI 359

Db 298 DISKPNMDSNYWEAAIFEALANEKLMNYPYKYMEQVPSEKERILSFVRN-HDINDL 356
Qy 360 FLPLDDIKVPLEV-----KIAFANNVINOALISLKDSYCSLDVINQINQINRYKILNDML 414
Db 357 ILPLGDIKISQLEILLRLKAATGKTFFNAFIISNDLSITLANNLSQLENRYEILNSII 416
Qy 415 NP--SINEGTD--FNTMKIFSKLASISNEDNMFMKITNTYLVKGVFPADPVESTINLSG 470
Db 417 QEFKICETVDSVINSVSELVLETTPKNLSMDGSSPVQIIGYLSGFKPEVNSTVFFSG 476
Qy 471 PGVVTGAYQDLMFKDN-----STNHLEPELRNPEFPKTKISQITEQETSLKSFNQ 524
Db 477 PNIYSSTATCTYFIKNTFDMLSQONQEIFEAS-NNLYFSK-----THDEKSSMLLRS 529
Qy 525 ARAKSEFEYKGVFEAGALCEDNDLPAONTVDKDVSKKI--LSSMKTRNKEYIHYIV 582
Db 530 NIASKEFKLIKVIIGTLYNEGLNPKWKVTTSELLKVIIEVNSTKIYENDLNMIL 589
Qy 583 QLQGDKISYEASCNLFKDPYSSILYQKNIIEG-SETAYVYVADAIEKEIDKVRIPVQIS 641
Db 590 QIQGDDI-SYBSAVNVFGKNPKSIL---IQGVDDFANVFYFENGIVQSDNINNLSRFN 645
Qy 642 NKNVILKTLFTGHGKSEBNTDTFANLVDVLSSEI-----ETILNLAADISPKVIEINLL 696
Db 646 DIKKIKLTLIGHGENVENPKLFGKTYNDLYTNIKPKQLQHLEREGVILKNKYLKINIL 705
Qy 697 GCMNFSYSISAEFYPGKLLKIKIDRVSELMP-SISODSITVSANOYEVRINEEGREIL 755
Db 706 GCYMTPTKVDINSTFVGKLFNKIS---RDLQPKGFSKNQLEISANKYATIRINEGKREVL 762
Qy 756 DHSGKWNKESIIKDISSEYISFNPKENKIIIVKSKYLHELSTLLOEIRNNANSSDIDL 815
Db 763 DYFGKWSNTDLIAEQISNKVYVWNEVENTLSARVEQLANKVAFAKDI-----NSLIQT 817
Qy 816 EKKVMTCEINVASNTDQIVEGRIBEAKNLTSDSI---NYTKNEFKLIESDSLSYD 871
Db 818 TNNQELKQSLVNTVADLTITLYSELLKEDIPFELDNQIKERIILNEISRLHDFSNIL 877
Qy 872 LKHQGLDDSHFISFEDISKTENGFRIRPFNKETGNSIFETEKEIIESEVATHISKISN 931
Db 878 FYQKNISNNMILFDSIIKEKYINVKLANKITGETSVIKTYSDSDSLWNFTNKYKIVDD 937
Qy 932 IKDITFDNVNGLKVKYNLDAHEVNTINSAFFQSLIEYNTTKESLSNLSVAMKVQVYA 991
Db 938 IKGIIVKDIINGEFIKADFEIQONPSLLNSAMQLLIDYKPYTEILTNMTSILKVQAYA 997
Qy 992 QLFSTGLNTITDASKVVELSTALDETIDLPTI-SEGLPIATIIDGVSLGAALKELSET 1051
Db 998 QIFQISICAIQEAETEIVTIIISDALNANFNILSKLVGSSVASVIIDGINLIAALTELKQV 1057
Qy 1052 NDPLRQIBIAKIGIMAVNLPAATAIVTSALG---IAGFSIILLVPLAGISAGIPSLVN 1108
Db 1058 KTNFERKLEIAKGVMSYIGFISLSSLSLGLGATAVSEILGVISVPVAGILVGLPSLVN 1117
Qy 1109 NELLIQDKATKVIDYFKHISLAETEGFTLDDKTIIMPQDDVLVLEIDFNNSITLKGCE 1168
Db 1118 NTLVLGEYNGIIDLVSFKYFPIVGNKPS--IQDNIIPYDDIAITAILETKTNKFKYGVAK 1176
Qy 1169 IWRAEGSGSHTLTDIDHFFSSPSITYRKPLWSIYDVNLNKKKEIDFSLKMLVLPNAPNR 1228
Db 1177 ISGFEGGGGHTTWGNIDHIFSPSLDHYIE-LSYFALKLNDTNLP-KGNVVLPSGLNK 1234
Qy 1229 VFGYEMGWTTPFRSLDNDGTLLDRIRDY-----EGQFYWRYFAFIADALITLKPYE 1283
Db 1235 VYKPEISAIAGANSQEGNGVEVLNIRNYVDSNGTKFPWKYEAPP-EYSFSYMRVEYF 1293
Qy 1284 DYNVRLNDGNTRSFIVPVTTEQIRKNLSYFVGGSGSYLSLSPYNNMIDNLAVENDT 1343
Db 1294 DTKANVILDNENKTLIIPVLTIDENRNKISYEILGDDGGQYNVILPVTQNTINIVSNKDI 1353
Qy 1344 WVIDVNVKNTIIESDEIQKGLIENTILSKLINEKIIILNNHTINFGVDINEENRFS 1403

Db 1354 WNPDSYIVKESKIEDNKFVLDGFINNIFSTLKVSDNGFKIGKQFIS-----IKNTPRAIN 1409
Qy 1404 LIFSILEDINIIIEIDL-VSKYKILLSGNCMKLIENSDD-IOOKIDHIGFNGEHQKIP 1461
Db 1410 LSGFINNNT-VIVSYIYNHEKNS:TISSDLNDIKNPDNLDNINYLIGLSISDNTIN 1468
Qy 1462 YSVIDNETKYNFIDYSKEGLFTABFNSIESIIRNYMPSDNNLIYSSKOLKDIRIINK 1521
Db 1469 CIVRNDVYWEKI-----FLNEK-KLVFIQNELEHLVDS-----VNK 1506
Qy 1522 GDVKLLIGN-----YFKDDMKVSLSTFIEDT-----NTIKLNGVYLDENGVAQI 1565
Db 1507 -DSQYILNPNVNVKYGKIVGFTLNSTENKYSLYIENKMKLGLYL-ESSVFKT 1564
Qy 1566 LKPMWAKALNT-SNSLMNFLESINIK---NIFYNLDPN---TBFILDTN---FIIGS 1616
Db 1565 IQDKIYSKEKVDYIILSLIKFTVNIQLCPFMIVSGVDENNRNLYEVLMTNKNKIING- 1623
Qy 1617 NSIQGELICDDKNTQPVFINKIKETSVLYVGNRQNLIVEPSVHLDDSGNISSTWIN 1676
Db 1624 ---GYWE---ND-----FNNTKIVDF-----EKNVIVSGGNKUNLSEGLDADT-ID 1662
Qy 1677 PSOKLYLGIDRYVNVKVIAPNLYT 1700
Db 1663 VLDKLENL---YDSVIIIPKVT 1684

RESULT 6
T00296
toxin B - Escherichia coli plasmid p0157
C:Species: Escherichia coli
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 17-Nov-2000
C:Accession: T00296; T42195
R:Makino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C.; Kubota, S.; Shinagawa, H.
DNA Ref. 5, 1-9, 1998
A:Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic E. coli strain O157:H7, substrain RIMD 050952
A:Reference number: Z14127; MUID:98290540; PMID:9628576
A:Accession: T00296
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-3169 <MAX>
A:Cross-references: EMBL:AB011549; NID:94589740; PIDN:BAA31815.1; PID:93337056
A:Experimental source: strain EHEC O157:H7, substrain RIMD 050952
R:Burland, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Sofia, H.J.; Blattner, F.R.
Nucleic Acids Res. 26, 4196-4204, 1998
A:Title: The complete DNA sequence and analysis of the large virulence plasmid of Escherichia coli strain O157:H7
A:Reference number: Z22068; MUID:98391744; PMID:9722640
A:Accession: T42195
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-236, 'D', 238-1887, 'I', 1889-3169 <BUR>
A:Cross-references: EMBL:AF074613; PIDN:AAC70163.1
A:Experimental source: strain EDL933; serotype O157:H7
C:Genetics:
A:Gene: toxB
A:Genome: plasmid p0157
A:Note: L7095
C:Superfamily: Escherichia coli probable cytotoxin
C:Keywords: cytotoxin

Query Match 5.9%; Score 507.5; DB 2; Length 3169;
Best Local Similarity 20.08; Pred. No. 1.9e-10;
Matches 420; Conservative 338; Mismatches 716; Indels 621; Gaps 105;

Qy 14 YVKPRI---QEDVEYVAINAL--EEVHNMSESSVWEKYLKDKDINNLTNYLTKYKSGR 68
Db 160 YIKRKTGAEDQTTTITQSLINELLNGVDRNTI-PFQKISELNDIIHSYENWQIKNSR 218
Qy 69 -----NKALKKFKB-----VLTMEVLEKKNLSLTPVKNLHFTWIGG 105
Db 219 KGIEILVKQELLSSLLNVNKNQKSDNASKIINLLIGIEYQSHKV-DIEPFIHAVVAG 277

Qy 106 QINDTAINYNQKVDNSDYTVKVFYDSNAF-----LINTLKT----- 144
Db 278 APPDNTFSYITAFINTYKDYTYLLWIDPNFAFGAAKFSGLKNIAMVAMRLARTINPHLA 337
Qy 145 -----IVESATNTNLE--SPREMLNDPEFY-----NKFYRKMELIYDQKHF 186
Db 338 EEMNEVILKIQNTIEFKETREBLKELERNYKSLTSEKPKFNFFLESIMIGMODNY 397
Qy 187 IDYKSOIENPEFI-----IDNIK--TVLSNEYSKOLEALNKYIEESLUNKITANNGN- 238
Db 398 FTYCISNGISNTDIOISLDFLTNVKLSPVQDNDFKSTVEKNKRDIDLLKNTISQKFGDR 457
Qy 239 ----DIRNLEKFADEDLVRLYNQELVERWNLAAASDILRISMLKEGQGVYLDVILPGIQ 294
Db 458 FOURDINTLESFKPKPDYFYQEMLLRWNYAASQVRINILKEYGGIYTDITDILPAYS 517
Qy 295 PDLFKSNKPSDITNTSWEMIKLEAIMKYK--BYIPG--YTSKNFOMLDEVOVSFESAL 350
Db 518 DKVSOIINE-KSDDKRPFEDLKLRIISIESILSIKGEKYSIKH-DGLDETTLNQLNNIL 575
Qy 351 SSKSDKSEIFLPLDDIKVSPLEVKIAPANNVI-----NQUALISLSDSY 394
Db 576 -SEIEK---LTIDDY--FKPVETKVRDTPKIFRYOKWTENTWNRGNFNMLTHGSK 629
Qy 395 CSDLVINQIKNRKYLNDNLNPSINEGTENTMTKIFSDKLASISN-----DN 443
Db 630 CIDFILSGQKKQYLEL-QRIDNISYNNLFYT-----EDLKSLLNVVAIGGIPAKKYLEH 683
Qy 444 MPMIKITNLYKVGFPADVRSTINLSGP----- 471
Db 684 GLF-----SEVRQDGTIPYVVSTNLISGPDIMRMQKVKYKSLGRIGBHVHKDNKLSDVNF 739
Qy 472 -GVYTGAYQDLLMP---KDNSTNIHLLEPE-----LRNPEPKTKISQIT---EQEIT 517
Db 740 LGTVASSNKNKSPNMLNPVSGINDITPDDESSWAVRNNDINKILPEKINCHVPEKLP 799
Qy 518 SLWFSNQAQAKSQEEYKVGFEALGDDNLPQNTVLDKQVSKILSSMKTRKEY 577
Db 800 SLY-----YEIDSRFFQGW---DNKSIKHVTEINKDLI--KDNLLTSSNID 843
Qy 578 IHYIVOLQGDKI SYEASCNLFSPKDPYSILYQKNIEGSETAYYYYVADAIEKIDKYRIP 637
Db 844 VKLIUKI--DRELYAISKI-----DNPLARSIRLQLOLANVYTSNTEPENTINFI 895
Qy 638 YQISNKNR-----IKLTFIGHGKSEFNTDTFANLVDYS-----LSSEIE 676
Db 896 YDFYKQKQDLSAIAKL-----FSRNDADTKIIVWNSVMEKNVFLREVIS 941
Qy 677 TILNAXADISPKYIEINLGCNMPYSISABETYPQKLLKIKDRVSELMPISQDSI- 735
Db 942 CVLRSKKVD---SYINEN-----KKNLSKEDA--GALRDYAKLKMKELFSLDDDDGYK 989
Qy 736 -TVSANQYEVRIINEEGKEILDHSGKWKNEESIIKDISKEYISFNP----- 782
Db 990 KIITTNAY---IKERDKL-----SGIYNIENSIISGHSFDIRSNOHEWGDISTVEQF 1041
Qy 783 KENKIIVKSKYLHELSTLLOBIRNANGSSDIDLEKKVMLTECEINVAS-----NIDRQIV 837
Db 1042 KKFEFYVKS--LSAKSIFDDIKNK-YITDPETKRVLYHQLDSDIKERIAFLDISHVAY 1099
Qy 838 EGRIEAKNLATSINIKNEFKLIESI-----SDSLYDLKHQNG 877
Db 1100 PGSLEKILQLS---GVVFSIDIIAEBVLLASVGVSHGVVYVPAPSDKLELLRRHT 1155
Qy 878 LDDSHFISFEDISKTENGFRIRFKETGNSI---FIETEKEISEVATHISKISINIKD 934
Db 1156 KSNSEWI--EKITP---YVYDILSDNVSVNVRPLPEEQKKILNDIKLEISKVS---E 1206
Qy 935 TIFDNNVCKYKVNLDAAHEVNTINSAPFTIQSLIEYNTTK-----ESLSNLSV--- 983
Db 1207 QYFMKLTQKSSVIGIKYSVDFDRYENLFLSLPINQMLTLPFMYRYFEMLYDIHIGIE 1266
Qy 984 --AMKVQVYQALFSTGLNTITDASKVVELV-----STAL-DETID 1020

```

Db 1267 NKANREITYKFSNLNLDPLNDRVNLGLIKKYKYLSEIHRHTLNSTSFADISIP 1326
QY 1021 LLPTLSBGLPIATII-----DGVSLGAIAK-----ELSETNDPLL 1056
Db 1327 LLQI--CPBITIINKTEYVGHQTNAMTVASWKPYDFNSLGNAINSDIKSVDPAL 1383
QY 1057 ROEB-AXIGMAVNLTAATAIVTSAL-----GIASGFSIL 1092
Db 1384 HTIVEQAKYNLWNDFYNTHASWDTIARQKSTNIEFHPQSLFDRDRSGKGLGSL 1443
QY 1093 LVPLAGISAGIPSLVNN-----ELLQDKATKVIDYFKHISLAETEGAFLLDDKIIMPD 1148
Db 1444 YLDTGGVGGYQKLRHNDITASTLYQK-----YNDNLKLSNRDFFLEKQRLITMSN 1497
QY 1149 DVLSEIDFNNSITLKGCEIWAEGSGHTLTDI--DHFFGSPSYTYKXPLSYDVL 1206
Db 1498 EL-----GNRLKNAQLEVLK--DPLTLEGILYQRRISLLITTEVHSLAQOIS 1547
QY 1207 NIKK-----EKIDFSKOLMVLNAPNRVFGEMGWTGPFSLNDNDGTLL 1251
Db 1548 SFWRVTDPNFGCHDFHSLAQAALTIFKNI-----TSNR-----NFSLYGSGI-- 1589
QY 1252 DRIRDH-EGOFYWEYFAFIADALITKLPRYEDTNVRLNDG--NTRSFIVPVIITEQ 1307
Db 1590 --VKIYFSESNNWKYIKPLVQGSLLRDYILTPPEKISTSGGSLNIMHLPV----- 1642
QY 1308 IRKNSLYSF-YGSGGSYSLSPYNNMIDNLVENDTWIDVNVNKNITIESEIQ-- 1363
Db 1643 -----SPYIDIGV-----INGNRISSEST--DVKNKIRSLKINGDILQHYI 1681
QY 1364 -----KGLIENILSKLIENKI-----ILN--NHTNFY 1392
Db 1682 NTHYLSBEQTQKIDIVDFLGTQDNTIKVKLESIDKIPSEIQQLPHLSILSRQKEHVNLL 1741
QY 1393 -GDINE-SNRP-----ISITFSLIEDINI-----IIEI--DLVSKSYKILLSGNCMK 1435
Db 1742 SGLLDEFNKLKQGLSLKTNVLSVNNPKESKINSIDTVEVTVDLQGLRYRVIDTRVIG 1801
QY 1436 LI--ENSSDIOQKIDHIFNGEHOQYIPYSVIDNETKNGFIDYSK-----KEGFTAFBS 1489
Db 1802 LTFKEGINSLSALEHNM-----IDAIMSVIGLVQVARIWKNNDNISADHA 1848
QY 1490 NE-SIIRNIYMPDSNNLFYSSKOLKIRIINKGDVK--LLIGNYFKDDMKVLSFTIEDT 1547
Db 1849 GAVSDIKNIV-----DKFLGGILTJNNRVNPGVSGASLEGFTSSGLEVCAS----- 1897
QY 1548 NTKLNGV--YLDENGVAQILKFNNNAKSAALNTSLSLWFLSNINIKNIFYNMLD----- 1600
Db 1898 --RMGGTAGRYL--SNVAKVIKL-----PLLDI--GINIWSLYDSSLNHAKA 1938
QY 1601 -FNIEPI--LDTNF-IISGNSIQGF-----LICDKDKNIQPYFINFKIETSVTLYV- 1650
Db 1939 TQIIEYISTADIVSFSSINTALSIGAIAYPPLAIA-----IVPITI-PSHEVKNYAVTVN 1992
QY 1651 --GNQNLIPEVSXHLDD-----SGNISSTVINFSQKLYG--IDRYANKVII 1694
Db 1993 QINERHKLWLEAKYLDNGSAKVLNINKATGIDLSNNQVLGNIVLDRENPPL 2047

```

RESULT 7
 T28676
 rhothry protein - Plasmodium yoelii (fragment)
 C:Species: Plasmodium yoelii
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
 C:Accession: T28676; A45521
 R:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
 Mol. Biochem. Parasitol. 76, 329-332, 1996
 A:Title: Comparison of two members of a multigene family coding for high-molecular mass
 A:Reference number: Z20507; MUID:97077455; PMID:8920022
 A:Accession: T28676
 A:Status: preliminary;
 A:Molecule type: DNA

A:Residues: 1-2401 <SIN>
 A:Cross-references: EMBL:U36927; NID:G1041784; PID:G1041785; PIDN:AA41263.1
 R:Keep, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
 Mol. Biochem. Parasitol. 42, 241-246, 1990
 A:Title: Identification of the gene for a Plasmodium yoelii rhothry protein. Multiple c
 A:Reference number: A45521; MUID:91101660; PMID:2270106
 A:Accession: A45521
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 2260-2401 <KEE>
 A:Cross-references: GB:M34281

Query Match 5.4%; Score 468.5; DB 2; Length 2401;
 Best Local Similarity 20.2%; Pred. No. 3.3e-09;
 Matches 410; Conservative 333; Mismatches 686; Indels 596; Gaps 105;

```

QY 6 KAQLQMYVYKFRIOEDYVAILNALAEYHN--MSESSVVEKYKYLKQDINNLTNLYNTY 63
Db 118 KDTMKKIILLIIQYM-NEFKGLNDAMTKLNGEGISQKFEVINNQIKQKFDKSTYDE- 171
QY 64 KXSGENKALKKPK--EYLTMEVL-ELKNSLTPTVEKNLHFIWGGQINDTAINYNQWKD 120
Db 172 KKEGFESSELEAKNWEKKLEIITELKKQNETVOLDIKIRELIKQIKD-----IIEQKI 227
QY 121 VN-----SDVTVKVYDVSNAFLINTLTKTIVESATNTNTLESFRENLDNDFDYNK 170
Db 228 VNDLKLKLNKKIKETIEKIEYKAV---DLKKEIKD--NVYID--ELAKEPPYQITK 279
QY 171 FYRKMEIITYDKQHFIDYKSIQIEE--NPFFII--DNI-----IKTYLSNEY 214
Db 280 YIEKKNEIYNTIKSDPDKIYVGDIEQLYNEMPSVVOESNIEHIEKTEILTKTKIDVNY 339
QY 215 S-----KDLKLNKYIEESLNTKANNG-----NDIRNLEK----- 245
Db 340 NNIQNMETETVKASHLKIETNNKLSSETLIDIIKYIYEITNELNKTLEDFKNKEGLNKK 399
QY 246 ---FADEDL-VLYNQELVE---RWNLAASDILRISMLKE-DGGVYLD--VDILPGIOP 295
Db 400 IDEYAKENQVLNYSKNIILEIKKHYN-----DOIINDNIKEKAKQNDVQFKEHKTIPP 454
QY 296 DLPKSIKNDPDSITNTSWEMIKLEAIMKYKEYIPGYTKNFMDLDBEVOFSFSSALSSKD 355
Db 455 NEMK-YQKP---STEIKMKDEFLSKNKY-----NDFDKYKKEVESEHNKFTELN 503
QY 356 KSEIFPLDDIKVPLEVKIAPANNSVINQALISLQSY-----C--SDL 398
Db 504 K--IKTEVSDEEIKKYENKFN-DSKSLINETKKSIEEYQNINTLKKVDDYIKVCLINTNE 560
QY 399 VINQIKRYKILNDLNPSINEGTDPNTTKIPFSDKLASI----- 438
Db 561 LITNCHNKQTLKDLKNQIKTIKETSIDKIYTDKFENILTDKKTELETKTGTGSLNHH 620
QY 439 -----SNEDNNMFM-----IKITNVLKVGFPADVSTINL 468
Db 621 ESNKKELLFYDLKANLKNKNEMLYKFNKEKAVEDIKKNV-----DINKIVSN 673
QY 469 SGPGVYTGAYQDILLMFKDNSTNTHLLELRLNFPFPKTKISQLTQEOEITSLMSFNOARAK 528
Db 674 IEITIYTSYN---INEDIENEIG-----KSELLANTKVLKVKVANVTNL----- 715
QY 529 SQFEYKKGFEFEGALGEDDNLDPQONTVL--DKDYVSKILSSMKT-----RNKEIYHI 581
Db 716 NEIKELKDYDFQDFGKKNIKYPDENKINDIDITLNQIKDSIETLTETIKKNS--NHI 773
QY 582 VQLQG--DKISYEASCNLSKDPYSSILYQKNIEGSETAYVYVADAETKEIDK--YRIP 637
Db 774 DEIKGQIDKLKVKPNKTMFNEDP-----KEIEKKLENIV 807
QY 638 YQISNKNIKLFTIGHGKSEFNTDTFANLDVDSLSSSEIETILNLAKADISPKYIEINL-- 695
Db 808 EKIDKKKI-----YKIDKLNLEISKIEN-DKTSLE-KLKNINLSY 847
QY 696 ---LGCNMF-----SYSISABETYPGKL-LLTKDRVSELMPSISQD----- 733

```


Db 553 KIEKIHASIVNTSEETEREIGINIESLNTKVEKV---KENVTLNLIKIKELKHVDFSD 609
Qy 504 -----PKTK-----ISQTEQITSLWSPNQARAKSQEYFKYFEGALGE 545
Db 610 FGEGNTKYTKIKINDIMAVSQIDQHINGLDDI-QKSESYYSEMKE-----Q 660
Qy 546 DDNLDPAQNTVLDKDYYSKILSSMKTRNKEYTHYVLOGDKISYEASCNLFSDKDPYSS 605
Db 661 INKLEKVSNTSEISNDNVE-----GKKKQOI-IVTKIDKKNIYEEINKLUSE----- 707
Qy 606 ILYQKNIGGETAYYYVADAIEKIDKYPYQISNKRNIKLTFIGHGKSEENTDTFAN 665
Db 708 ---ISKLEKNTS-----LEKVKDNLVSG-QNLGNLFLEQIDEEKKAE-NTIKS 753
Qy 666 L-----DVSLS-----SSRIETILAKADISPKYIEINLGNMPSYISABEYTPGKLL 717
Db 754 MEAVIDDLNLIKKSQIEITEMDI-KWDINKE-----MEALKISHDDO-----K 796
Qy 718 KIKRVSELMPISQDSITVSANQYVRINEEGKREILDHSGKWIN--KEESIIKDISK 775
Db 797 KCHDK-----SKN-----HKENISDIYDKSKIIQDPSRESINDIKK 835
Qy 776 EYISFPNKEN-----KIIIVKSK-YLHELSTLLQEIENNAN 809
Db 836 LQKNVSEQNHNSDINOCLNEVANYNILKNIKLIIDKVKETSEIKNKNINDELN 895
Qy 810 SSDIDLEK---KVMLTECEINVASNIDRQIVEGRIE-----AKNLTSDSINYIKNEFK-- 860
Db 896 NSEKVIKIEGDLSEKCRSKINSTLDDKIDCEIKINVLKKNILNEETN-ITNHPKNA 954
Qy 861 -----LIESISDSYDLKHQNGLDDSHFISPEDISK---TENGFRTFIN 902
Db 955 BEYNKIVLSNFFNEMADNKSQVILEIKKNGTND-HDYNIKELKSHKDKSNGYK---T 1009
Qy 903 KETGNSIETEKELFESYATHI-----SKESINIKOTIFPNVAGKLVKKNVDAH 954
Db 1010 EADQKKAQIKQNKELFEQYKEEVTLLNKYAVELKKNFKDT-KNDSKQIIKEIK--DAH 1066
Qy 955 EWTNLSAFFIQL-----TEYNTTESLGNLSVAMQVQYQALFSTGLNTITDAS 1005
Db 1067 NYCTLESKSEKKWNEIKNEKIHIDEVANNDKNKAIT-SIKVSEVPFKT---KIKIN 1122
Qy 1006 KWVELVSTALDETIDLLPTLSEGLPIIATIIDGVSLGAAIKELSETNDPLRQ----- 1058
Db 1123 EIRTKSDCLKETNDLEQIS-NLSIDTQETKLTEGKQLTKLELLESLEKQKKNIEDQ 1181
Qy 1059 -----EIEAKIGIMAVNLTAASTALVTSALGIASGFSILLVPLAGISAGIPSLVNNELIL 1113
Db 1182 KXELDEVANSKI-----KNIENTYN----- 1200
Qy 1114 QDKATKVIDYFKHIS-LAETEGAFLLDDKLIMPQDDILVLEIDFNNSI-----TLG 1165
Db 1201 QHKQNYEIGIVEKINEIAKTNKQIESTKELIKPTIOHIS--SFNANDLEGIDSDENLG 1258
Qy 1166 KCEIWEAEGGS-----GHTLTDIDHFFSPTSITRKPWLSIYDVNLINIKKEIDFSKD 1218
Db 1259 K---YNTENGNIYEFKISYNLITNYLETVSKESITYNQ-----IQNKRIDTQKE 1305
Qy 1219 LAMVLNAPNRVFGYEMGWTGPFRLSDNGTKLLDIRDHYEGQFVWRYFAFIAALITKL 1278
Db 1306 L--LKNIN-----VNKAQSYLDVIK---ENEF-----DRIVTHF 1335
Qy 1279 KPRYEDTNVRINLDGNTRSFIVPVTTEQIRKNLSYSFVSGS-GYSLSLSPYNNNIDLN 1337
Db 1336 KKKLNTVNDNF-----KN-EYSKVNNEGFDNISINTKVNSTDEN 1374
Qy 1338 LVEN-----DTWVIDDVNVKNITTESDEIKGELIENILSKLNIEDNKIILNHTINF 1391
Db 1375 SLNLINLOQKEMANIVNTYYSKYEAENIFRN--IPKLANLIK-----IKNSGIDL 1428
Qy 1392 YGDINESNRFISLTFSLIEDINIIL-----EIDLVSYSKILLSGNCKMLIENSDDI 1443

Db 1429 SKDKIA--ILSYLDSKTEDTLFIPSPQKKTETTYKISDSYSLILD-----ILKKSQEL 1481
Qy 1444 QOKIDHIGFNGHQ--KIPIYSYIDNETKNGFYDYSKKEG-LFTAESFNSIIRNTYMP 1500
Db 1482 QKK-----EQTKLI---PENRRLEKVGQATNELRGTLSDLYKKEKILSEV--- 1526
Qy 1501 DSNLFIYSSKLDKI-----RIING---DVKLLIGNYPKDDMKVLSLFTIEDTNT 1549
Db 1527 ---KLLLHKSNELNKLSNCFONDTILESSEYQDVKEKSNNYKQEKLGIDFNVTDMEE 1583
Qy 1550 IKLNGVYLDENGVAQILKFMNNAKSALNTSNLSMNPLESINIKNI FNNLDPNTEFILT 1609
Db 1584 -KFN-----NDIKVIBELENNYDSBENNILOSKOKLKTNFNAEKIKDDKILIEK 1636
Qy 1610 NFITSGSNSIGOFELICDKNIQPY-----FINFKIKETSYTLYVGNRQNLIVEPSYHL 1664
Db 1637 NDLI-----DKLIETRKNCMLFTHTLAEFLKIKITDYSKFIESATKFSKFLKYI 1687
Qy 1665 DDSGN-ISSTVINFSQKY-LYGDIDRYV-NKVIIAPN 1697
Db 1688 GDTNSLNDIATLQLKYDLHQINKYVTSKLSDATN 1723
RESULT 9
F82885
hypothetical protein UU482 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: F82885
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Caswell, G.H.
submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mi
A:Reference number: A82870
A:Accession: F82885
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4688 <GLA>
A:Cross-references: GB:AE002145; GB:AF222894; NID:G6899476; PIDN:AAF30894.1; GSPDB:GN00
A:Experimental source: serovar 3; biovar 1
C:Genetics:
A:Gene: UU482
A:Genetic code: SGC3
Query Match 5.0%; Score 431.5; DB 2; Length 4688;
Best Local Similarity 19.1%; Pred. No. 1.7e-07;
Matches 420; Conservative 329; Mismatches 743; Indels 709; Gaps 103;
Qy 13 VYVKFRIODEVVAIINALE-EYHMSSESVVVEKYKLDKINLTDNVLNT---YKSGR 68
Db 403 LYEVIGIKKTNEVDVNLKQIPYHNS-----INNLSNALNTPYQYTKNGD 449
Qy 69 NKALKKFKELTME-----VLELKNS-----LTPVEKNLHFIWGGQINDTAINYNQWKD 120
Db 450 INLIAKVAFYVYVQVYGFKDQNNQEHQILAKVKKDGTAFTGTALNNSNYSLDKIVS 509
Qy 121 VNSDYTVKVFYDSNAPLINTLKTIVESATNLTLSFR-----ENL----- 161
Db 510 VSNPQNLV---SNFDLSKQKLIKPAANASVSTKTKTQILENLNLDLNLQKLVATVP 566
Qy 162 --NDPFDYNNKFKRMEIYDKQ---KHFDYKYSQIEEN----- 197
Db 567 DNNDKEYKVVANVDQNNKIIFSDNLPKGYI-YHLAKVENNDLNKVINLNDPELKDKIID 625
Qy 198 -----PEFIIDN-----IKTYLSNEYSKDL--EALNKY-----IBESLN 230
Db 626 KRLNLLDSDPDTYDNGNLEHTQLANDLDDLQKALNNANVKGIVVDQDQIEHEID 685
Qy 231 KITANNNDIRNLEKFADEBLVR--LYN-QELVERNLAAASDILRISMLKEDGGYLDV 287
Db 686 VSDANGKVIPTKLANNDPTKPNITVTLKVVYKQNNQPNIDLI SEEQLSGDNHISFK- 744
Qy 288 DILPGQDPDFKSIKDPD-----SITTSNEMIKLEALMKYKEYIPGVTSKNFMDLDEEVQ 343

745 Db -----KPTTAKTKENDDYEIFSFPNSPLANKKIKLTKETDD-----NNTNTKTVEAS 791
 344 QY RFSEALSKSDKSIFLP-----LDLIKV-----SPL----- 371
 792 Db IGLDGAIFKTSDDAIFAPDHRKYTLTKJEADNKKVANIDEISPLDIRVYNKQNGVADN 851
 372 QY -----EVKIAFANNSVI-----NOALISLSDSYCSDLVINQIKRY-----KI-- 409
 852 Db KHEFKIPDQKNKOLFAVYKDNKNNHIVPIKTDGKGVIVPNNNLPFPNKIYDFDKIVD 911
 410 QY LNDNLNPSINEGDFNTMKIPSD-----KLASISN--EDNMFMKIKITNYLKV 457
 912 Db LNEYFNKTLIDRNSINKDVSALNDGVDQARKLVKAPAVSNVTVNAINFQVNLFDNLKLS 971
 458 QY -----FAPDVRSTINISLGGPVYTGAVQ-----DILMFQDN 487
 972 Db YNRQFALTIKAVNNDLTKYATYDPPKNTNKKLNFDFETHLDANTKYKVDVLELNNK 1031
 488 QY STNIHLLEPBLNFPFKTKISQLEQETISLMSFNQARAKSQ-----FEYKGYPEG 541
 1032 Db EKPIKLIKDDVLNFEFTSSATTIN-----PPIWTKFDVVVKTNNDTTITFEIDDK---DN 1084
 542 QY ALGEDDNLDFQAVNTLVKDYKSKILSSMKTKNKEYIHYVQLQDKKISYEASCNLFSD 601
 1085 Db ILKNDOKI-YAQUALMDDLLDVTDVINPLVNFST-----NKIASINGLDLKGNS 1133
 602 QY PYSSILYQKNIEGSETAYYYYVADAEIK-----EIDKTRIPYQISNKNRNIKLTFIGHGK 655
 1134 Db KYSI-----XNL-----YLNQDNQKVYLFKNDVTKYE-QHFTNPKINLSF---NK 1177
 656 QY SEBNTTTP---ANL-----DVDSLSEIETILNLAKADISPKY 690
 1178 Db SAVEQDIFADHANLFDIDYDQKLRIINEDVKIYYQNDITKNELQ--IGYKV-VANNK 1234
 691 QY IEINLKGCMFNSYSIAEETYPKLLKIKDRVSELMPS-----ISDSITV 737
 1235 Db IKFNLVG-----LAKETTYVKKLEALNKGASSIVNSEFOLLDPNTVFNSTNKNITLV 1287
 738 QY SANQY-----EVRINEEGREILDHSGKI-NKEESIICKDISSEY 777
 1288 Db GLNSIDWANDQPIINAKINIGDDFQDNQOVKLIIVSNDNKEIKSAVTLIKGQRYQV 1347
 778 QY ISFPNKENKIIVSKY-----LHELSTLQEI-----RNNA-----NSSDIDLEKKVWL 821
 1348 Db EFSNLIKNRLYTSKIVYETNNQTLHLKDLTHQFSINPNSNAVSLKNTNIEITKRIIV 1407
 822 QY TECEINVASNIDQIVIEGRIE--EAKNL--TSDSINYIKNEPKLIESISDSLYDLKHQNG 877
 1408 Db -----NNDQSLISAKIEVDVDDIDNVLNTADKPNIVYQ-----LENSDANNLKLATN 1454
 878 QY L-----DDSHFISFEDI-----SKTENGFRIRFPINKETGNSIFTETEKE 916
 1455 Db AQVVVNNKKFLFDVLGLKINQYVIKEISFNSKPTNAY-FNFTNKNNNIIVSYDEON 1513
 917 QY IFSEYATHISKEISNTKIDTFDNNV-----GKLVKKNLDAHAHEVNTL----- 959
 1514 Db -----KISLSNNINPTSKYKPOINKNDKSNVNDVLDQVDKQLLANQYLR 1557
 960 QY -----NSAFIQSLIEYNTTKES--LNSLVAMKVQVYALQFS-TGLNITITDASKVV 1008
 1558 Db LKLQLNDNKTWTDPIFNNAKISFKLSNL-----IHNRAVELEGILYFDDQNSVN 1610
 1009 QY ELVSTALDEIDILPLISEGLPIATIDGVSIGAAIKELS-----EFTNDLLRQ 1058
 1611 Db DMTNNQISFNSKHKPKIEPEPSLTTI--NYDTNNAIKTVSAHNAQVHFKLTKNTDEALSN 1668
 1059 QY E-----TEAKIGIMAVNLTAATAIVTSALGIASGFSILLVPLAGIS 1100
 1669 Db QQVEAVFAPTNLNDQKVVEAKLNNVTSNFESELEFNSLGRSEETVELIKVTKNKP 1728
 1101 QY AGIPSLVNN-----ELILQDKATKVIDYFKHISIAETEGAFTLDD----- 1141
 1729 Db NKAYELLKNGVIFPEYKNGSGAVEFTTQKFEHKVIDVWSSTSTNTTQEIETKIDIGIORA 1788

1142 QY -----KIIMPQDDDLVISE-----IDFNNSNI-----TLGKCEI 1169
 1789 Db WNNKKLELYVESNILGDPKIKTIVDNNNSVHLSFDPKKEYNLVLNNLKPGRYSLKKINI 1848
 1170 QY WRAEGSGHTLDDI--DHFF-----SSP9EITVRKPLWSLYDVVNLKKEKIDFS 1216
 1849 Db KEVDNGQDHEFVKEINVNNNSFVNLQSEITASSVEEINDRAP-----DKLNTQTTIKINLK 1903
 1217 QY KDLMLVFNAPNRVFGVEMGWTGFRSLDNDGKFLDRI-RDHVEGOFYWRYPAFIADALI 1275
 1904 Db DENDILK-----TNDIATYDNEQKVDALVKTNAQNKY-----LEAII 1943
 1276 QY TKLPRVEDTNVRINDGNTRSPIVFVITEQIKNLSYFSGSGSYSLSPYNNNI- 1334
 1944 Db TNL-----VFNKNVILKTEFKNLSQTF-----IKVKNTNVI 1977
 1335 QY DLN--LVENDTWI-----DVP--NVVKN-----ITIESDEIQKGLIENILSKUN 1376
 1978 Db YDESNLKLINNDFOIIGPLSTDVNSTQNVANNKHVLSSTLDFKVNPHISKNLAKFLK 2037
 1377 QY IE-----DNKIILNN--HTINPYGDINESNRFISLT-PSILEDINIIIEIDL 1420
 2038 Db FQNINGEVVYSPILTNSSIIIVNNKNVINFITLNLAKSNQLYRLVDVYVYDDNN--DT 2092
 1421 QY VSKSYKILLSCMKLIE-----NSSDIQOKIDH--- 1449
 2093 Db INDKNKVPKANNVTRIIDAPGKTTISKNNNTWNTSTSSQFEFVINSDDGNEVLNLEA 2152
 1450 QY IGFNGEHOQYIPVSYDNETKNGFDYSKKEGLFTABFSPNESIIRNIYPPSNNLFIY 1508
 2153 Db TISFKKGQOTLLTPKV--NIKQNN--KYLKQIINLEPENRYVLESILLAKPN--- 2203
 1509 QY SSKDLKDRIIRINKGV--KLLIGNYFKDMKVSLSFTIEDTNTI--KLVGYLDENGVAQ 1564
 2204 Db KTKKPLVVEILNKDDISFOAGYKVIQIKSQNPSTVDTKQRIKLDGIONAWNEKQL 2263
 1565 QY ILKPMNAKSAALNTSGLMFLSINIKNF-YNNLDPN-----IEFILDIN----- 1610
 2264 Db EITYSAND-----NSTKTAAIKLEKNKLEVEFELTNLEKNNTYFTTKIELINDNNTKTPFN 2319
 1611 QY FIISGNSIGQFEL--ICDKDN-IQPYFINFKIKETSIVLVGNRQNLIVE 1659
 2320 Db KDSIQDKFVLSNNQVGVNIIIEIQDRVHNLNSAKIRFELNDLNLV--SNDEQATIT 2377
 1660 QY PSYHLDSGNSISSTVNFISQKLYGI-----DRYVNVK 1692
 2378 Db YNNQTTISAKV--ITQNKQYLEATFSLNLVNLKNTIINKI 2415

RESULT 10

G70163
 hypotheical prctein BB0512 - Lyme disease spirochete
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
 C:Accession: G70163
 R:Fraser, C.M.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Whil-
 son, D.; Peterson, J.; Karavague, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugi-
 ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A>Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:98065943; PMID:9403685
 A:Accession: G70163
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-2166 <KLE>
 A:Cross-references: GB:A3001153; GB:AE000783; NID:G2688419; PIDN:AAC66876.1; PID:G2688,
 A:Experimental source: strain B31

Query Match 4.9%; Score 422; DB 2; Length 2166;
 Best Local Similarity 20.2%; Pred. No. 1.3e-07;
 Matches 435; Conservative 333; Mismatches 677; Indels 704; Gaps 112;

RESULT 11
T18427
hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18427
R:Lawson, D.; Bowman, S.; Bartell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18427
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3724 <LAW>
A:Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325379; PIDN:CAB11104.1
C:Genetics:
A:Introns: 307/1; 1545/2
A:Note: C0335c

Query Match 4.6%; Score 399; DB 2; Length 3724;
Best Local Similarity 19.7%; Pred. No. 1.9e-06;
Matches 402; Conservative 306; Mismatches 611; Indels 724; Gaps 109;

QY 20 QDEYYVAINA-----LEEYHNSSESVVEKYLKLDINNLDNYLNTYKKS-----G 67
DB 618 EDEYHVLKENLKEDANEYNDKEN---KNNKTEILK-SKYLENEKRTLEELKLRG 672

QY 68 RNKALKKFKELTWEVLEKNSLTPVEKNLHPFWIGGQINDPAINYNQWKDVNSDYTV 127
DB 673 KNIIFKDEKY-----NSLGEV-----IINEIQINEENKINDIQOG--- 708

QY 128 KVFYDSNAFLNLTAKTIVESA--TNTLESFRENLDPEFDYKNYKREMEIIVDKQKH 185
DB 709 -----NISQKIIQSSRTNDTFNIKDISLND--DLEKEKRKK-----KSOH 748

QY 186 FIDYV---KSQLEENPEFTIDNIKTY-----LSNEYSKDL-----EALNKYIEBSLKI 232
DB 749 FIDNLVKADKNEISENINKICDNNINNYDESINNYDESINNYDESINNYDESINNI 808

QY 233 TANGNDI--RNLEKPADEDLVLYNQ-----ELVERNLAAASDILRISMLKEDGG 282
DB 809 YDENINNYDENINNYDENINNYDEGINIKICDDNILENKNLKTINDIVQV-----EENN 864

QY 283 VYLDVILPGIPLFKSINKPSITNTSWEMIKLEAMKYKEYIPGYTSKNFMDLEEV 342
DB 865 ESIE-----KNELMISLNDK--INNTYMFK-----ENVDFINKI 898

QY 343 QRSPESALSQSKSEIFPLDDIKVSPLEVKAFANNSVINQALISLQSKDSCLDVING 402
DB 899 KR--ESLL--KIDK-----NIK-----DNN-----NDDEVIMDNFYENDFIIN- 932

QY 403 IKRKYILNLANP-SINEGTDFTNTMKIFSDKLAS---ISNEDNMWF-MIKITNYLKV 457
DB 933 --HMEITNKELDPLEINTQNEFTENLDIKKKYITNDHPFNADAKMFYEMNKILN----- 985

QY 458 FAPVRSITNLGSGVYTGAYQDILLMPKDNSTNIHLEPELRFEPFKTKISOLTEQET 517
DB 986 --KDMKKN-----KEGEFTTDTGSLQSHKI KKYNGEEKHDKNNNEEKN 1030

QY 518 SLWFSNQARAKSQPEEYKGYFEGALGED-DNLDFQANTVLDKDYV-----SKIL-- 567
DB 1031 ILYDENQV-----YSVLYSDHKIEQDIQDIHSIQTNICDENBEQINEENSKEGVRI 1082

QY 568 --SMKYTKNEYIHYIVQLQD---KLSYEASCNLFPSKDPYSSILYQKNIEGSETAYYY 622
DB 1083 SGTDMENKNMENKNDMEKNDMEKNDIEKNDMEKND-----MEKN----- 1127

QY 623 VADAEIK-EIDKYRIPQISNKNIKLTFIGHGKSEFNTDTFANLDVLSLSSEIETILN- 680
DB 1128 --DMEKNDMEKN---DMENENNE-----NKSDEIENKQNEYKNKNKNKNKNKN 1176

QY 681 -----LAKADISPKYIEINLLGCMFYSYSABETYPGKLL---LKIDKRVSELMPSTS 731
DB 1177 TIENNMIENNMIENNMIENNMIENNMIENNMIENNMIENNMIENNMIENNMIENNMIENN 1231

QY 732 QDSITV---SANQVEVRINEEGREILDSHGKWKINEESI-KD-----ISKEY 777
DB 1232 KWKIKIEDDAKQINQVNNQNKVHYKDKNDLNNVD-IINDVLSKDDKFNWINSKE- 1289

QY 778 ISFNPENKIIIVSK-----YLHELSTLLQEIIRNANSSDIDLEKQWMLTECEINVASNID 833
DB 1290 --INIKDFKVDKNDICLNLNELNK--QSYNGNTTEKCIDIH-----DFDYVENKD 1337

QY 834 QVIEGRIEAKNLTSPSINYKNEFKLIESDSLYDLAKHQGLDSDSHFISFEDISKTE 893
DB 1338 ---VHDKIHEDRKEYCD-----ESKLKYLPAADDIKOMR-----SFKINKKSKRE 1379

QY 894 NGPPIRINKETG---NSIPIE-TEKEIFSEYATHISKE-ISNIKOTIPDNV-NGKLIV-K 946
DB 1380 NFLIISYNEEIPFLIKLKVLEIIEPSIILKNKGHIKQFLPLGKYINDNLKGEKCISS 1439

QY 947 KVNLDAAHE-VNTLNSA--FFIOSLIEYNTTK-----ESLSNLSVAMKVQVVAQLPSTG 997
DB 1440 KIIIDLMLEYMTIISKXNGVLYFNNNYLQHLDFELHKLGLSTIKYINNYVHFLNIF 1499

QY 998 LNTTIDASKVVELVSTALDETIDLLPTLSEGL-PIATII---DGVSLGAA----- 1044
DB 1500 LKAVND-QKQVNIYDILLDDQKIIVSNPFGMPFKFLILKKNHIVSIHSGKHFNPOKIY 1558

QY 1045 IKELSETNDPLL-----ROEIEAKIGIMAVNLTAATAIVTSALGIASGFSILLVPLAG 1098
DB 1559 RRLLSKFNDDKOIDSNNKVEVEEK----- 1593

QY 1099 ISAGIPSVNNELILOQKATKVI-----DYFKHISLAETGAFTL-- 1138
DB 1594 -----HSELPANDKNEEKYKNVKNQENQENDKSVCSDEYVLSIS---SEKYNVYG 1631

QY 1139 ----- 1138
DB 1632 EDEQEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQ 1691

QY 1139 -----LDDKIIMPQDLVLS-EIDFNNSI-TLGKEIWRAGGSGHT 1179
DB 1692 DDDDDDDNDDDDDDDDHYDXPSDATQNTSINIKSIEEINNYENNTKNKDNKHT 1751

QY 1180 -----LTDDIDHFF--SPSITYRKPWLSIYDVLNKK-EKIDFSKOLMVL----- 1222
DB 1752 DESYVNLNKKNEINTYKCTSPDKYK---NYN--NLKKYCDMEHEKDLMLKHLV 1805

QY 1223 -----NPAENRVFGYEMGTPGFRSLDNDGTLLDR-----IRDHYEQGYWR 1265
DB 1806 LLKDEKENIKNNEGEE-----SIDN---IILDKNMSDQEVNENKDLNNEY-- 1852

QY 1266 YFAFIADALITKLKRYEDTNVRINLGNTRSFIVPVITTEQIRKNLSYFYSGGGSYL 1325
DB 1853 -----IENLSEHLAKHYDIEMIPIDIVGNIR--LYKLFIYNNMIKN--YSYLKNNFP-- 1901

QY 1326 SLSPYNNMIDNLVENDTWIDV-----DNVKNITIESEDEIQKGLIENILSKLNIED 1379
DB 1902 ---PYKKGKRGKAKGMLMINIKRHEYNLMKK---KNGNEKTG---ANNKKNKND 1952

QY 1380 NKIIANHTINFYGDINESNRFISLTSILEDINIITEIDLVSKSYKIL---LSCNCKWL 1436
DB 1953 KNDYINNNNNNNNNNNNDNN-----INV-----GYKMQNNINPNLYE 1991

QY 1437 IENSSDIOCKIDHIGFNGEHQ-----KYIPYSYIDNETKYNQFIDYSKKEGLTAFBSN 1490
DB 1992 IKNN-----FNLYHYWSLFYKCPVSYIND---NKIEGKK--FCVYVKN 2034

QY 1491 ESIRNIYTPSNLFI-----YSKOLKDI-----RIKNGDKVLL----- 1527
DB 2035 K-----YMLNNDHYMBFLKHPKPYAKSKVSKSVCSLVYFFNMDIQKKIILLETYE 2089

QY 1528 -----IGNYP-----KDDMKYS-----LSFTIEDTNTIKL----- 1553
DB 2090 KLIIHIDIGKYQYFFLSHQDKQKHKNMKNNDNDDVDNKNIEDAHIFEIHRIEGKTIS 2149

Db 1282 RNLANEINNKKQNDWCINEMNKKYGDICVYEMSLTYKETSLSKYEQKVYKDEKCS 1341
Qy 1481 -----EGLTAEFENE--SIIRNIYMPDSNNLPIYSSKOLKDIRIINKGDVKKLLI 1528
Db 1342 QYDEIRFOYDEKCFQYDEINKKYGALLNI--NITNKWVDSKVDRNNNEIISVDNKVEGI 1398
Qy 1529 GNYFK-----DDMKVLSFT--IEDTNTKLGNYVDENVGAQIILKPM--- 1569
Db 1399 ANYLKOIPELNEEIRLKEINKISLLYSNELNEKNSYDINMKHQEQ-----LLFLEKT 1453
Qy 1570 ---NNAKSALNTNSLMN-----FLESINI-KNIFYNLLDPN---IEFILD 1608
Db 1454 NKNEBEKIINLTASQSDAYKKSDESKLCAQFVDDVNIYGNISNNIRTNEYKYEEMPD 1513
Qy 1609 TNFIISGNSISGOFELICDKX--NIQPYFINFKIKET-----SYTLVGNRONLIVE 1659
Db 1514 TNIEERKNMELSKYIHLEENPRCKMIYIYENENIKSSNKIIGLYNRYRYGLREDLCKE 1573
Qy 1660 ---PSYHLDDSGNISSTVINFSQKLYGIDRYVNVKVII 1694
Db 1574 EIVPS-----KIGNISKNENNNKN-NTCDGVDEKVTI 1606
RESULT 13
D81702
adherence factor TC0439 [imported] - Chlamydia muridarum (strain Nigg)
C;Species: Chlamydia muridarum, Chlamydia trachomatis McPn
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C;Accession: D81702
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis McPn and Chlamydia pneumoniae AR39.
A;Reference number: A81500; MUID:20150255; PMID:10684935
A;Accession: D81702
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-3225 <RET>
A;Cross-references: GB:AE002312; GB:AE002160; NID:g7190482; PIDN:AAF39293.1; PID:g719048
A;Experimental source: strain Nigg (McPn)
C;Genetics:
A;Gene: TC0439
Query Match 4.5%; Score 391; DB 2; Length 3225;
Best Local Similarity 19.2%; Pred. No. 3e-06;
Matches 374; Conservative 306; Mismatches 670; Indels 598; Gaps 97;
Qy 2 NLVNAQOLQWYVKFRIQDEYVAIL-NALBEEVHNS-----SSVVEKYLKLDINN 54
Db 196 NILEKUSLTO-----EQQIKHNDNISLNKEAINRYSDLNRRKNSKGOSLLVRQAKILD--B 249
Qy 55 LTDNLYNTYKSGRNKALKKPFVLTWEVLELKNNSLTPVEKNLHPITWIGGQINDTAINY 114
Db 250 ILSTKTEREASNSVMTTIKGFTHRV-----PVEKNIHGIWAGSPPEGTDEY 300
Qy 115 INQWKNVSNYTVKVFYDSNAP-----LINTLKKTIIVESANNVTLESFRENLDNPEFDYK 170
Db 301 IKLFLHTYPPFSFLFWDKTAYGAAKFSSTLKRIAFDAVN-----SLREATPEPVKQFVQ 356
Qy 171 FYRGRMIIYDKQHFIDYKSOIEENPEPIIDNIITYLSNYSKDL----- 218
Db 357 RYDR-LKKSVDTSRDFDE--KQLSEQLVELYDNY-----NRFSEKIQSNFVLLHEM 407
Qy 219 -----BALNKY-----IEESLNKI 232
Db 408 ITIQDSFNYQLKGVGAITDETREIVLEKVLKVEEDLSHYKETIKKNKESIEKLVEI 467
Qy 233 TANNG-----NDIRNLEKFADEDLRLYN--QBLVERWNLAASDILRISMLKEDGGVY 284
Db 468 NDSTGRREVVKIDRLKSL--QDLTNSYVETEMLLRWNYAATDQLRYMLKEYGGIY 525
Qy 285 LDVDILPGIQDPLPKSIN-----KPDSTINT-SWEMIKL-----BAIMKYKEYIPGYT 331

Db 526 TDLIDIMPOYSQDLVQIMDVGGSRFFEHDKRRTLTSLFAALKLGSQKQTTVSFEAKKAMT 585
Qy 332 SKNPDMLDE-EVORSFESALSKSDKSEIPLPLDDIKV--SPLEVKI-----AFANNSVI 383
Db 586 LPFTLQDKSQISEIKYLETETQAKKSLFQPMQMDVTTVRDFMPILQRYHKWQTMVVRGL 645
Qy 384 NQALISLSDSYCDIYVINOIKRYKTLNDLNPASINEGTDENTMTKIFSDKLKASISNEDN 443
Db 646 NGLMMAKOSAVVDVIARQAAAYDEM-ALRONVVSGBEFFREL-----GDLEHVREKN 699
Qy 444 M-MFMIKITNYL-----KVGFAPDVRSTINLSPGVYITGAYQDILLM--FKDNSTNI 491
Db 700 IGGVLAK--NYLGSULFFDFRQDSVPVIGALSTIGISP-----DIIMDTMSDYFTNL 749
Qy 492 -----HLLPELERN-----FEPPKTKISQLT-----EQBIT-----SLMSFNQA 525
Db 750 GPVGEDFLYEGKLGKAFLGAYQAQKTPKGLTYVWLHPLSIGANDVPADASTWCETRQ 809
Qy 526 RAKSQF-----EYKKGIFEGALGEDP-----NLDFAQ-----NTVLDK 559
Db 810 HCAAEALLSDSISSEDEHPKGIIRERNVNPDSFKLWSKEAQIILSSDFADLLPRFNLLIES 869
Qy 560 DYVSKKILSSMTKNTKEYIHVIYVLOGDKISYEASCNLFPSKDPVSSILYQKNIEGSETAY 619
Db 870 SALDIHTLSAL--DRDIQHLFTKVQ-----KDPVASVA----- 900
Qy 620 YYYVADAEIKDKYR-IPYQIGN-----KRNKILTFIIGHKSEF----- 658
Db 901 ---VPSLQQLAEMIRAIPEIRNOVHILPEAQAHFADWKALQLYLHSHQPTEVVIWY 957
Qy 659 ---NTDTFANLDVDSLSSEIETILNLAKADISP-----KYIEINLLGC-----NMF 701
Db 958 SSTHTQIVFGKDLLAVERAAAKSL-MSDHPDSLITSYLKYKTQSHGLVTEFDQDFFF 1016
Qy 702 SYSIS-ABEYTPGKLLIKIDRVSELMPSISQDSITVSANQYEVRIINEEGREILDSH-G 759
Db 1017 ELMVDIAEPELHKQLLKIEQVN-----SGLYSHVEHSLG 1052
Qy 760 KW--INKEE-----SIKDISKEYISFNPKNKIIVKSKY-----LHEL 797
Db 1053 EWLKLSKEERKSKFLKILKETFQEBEEDSOQKHFWFEELYEKRHOVRKDPKAKTQEL 1112
Qy 798 STLLQERNNANSDDID-----LEKKVMLTECEINVASNIDRQIVEGRIEAKNUTSD 850
Db 1113 ITVTCB--SQRYAQADIDITYPAHKPFYQDLMKDGVAFEDISVITKYL-----LASD 1161
Qy 851 SINVIKNEFKLIESISDLYD-LKHQGLDDSHFISFEDISKTEGNGFRIRF--INKETGN 907
Db 1162 GVSGIITTDPIFPFPPSKQLIDAMKQSLG-----EDFGEHLYTLQWVYDWLSKET-N 1211
Qy 908 SIFITEK-----EIFSEYATH-----ISKEISNIKDTIFDNVNGKLVKKNLDA 953
Db 1212 SVTSEQAQKLFQKLHEKLEGYTHDLLIPIDGVSAL--GLRFSTEBEGKVSRLVTSIA 1270
Qy 954 HEVNTLNSAFIQS-----LIEYNTTKESLNSLVAMKVQVYAQLPSTGLNTITDASK 1006
Db 1271 PGVP--NSASVAMTSYLYGLFLITKDIOGSLTHEILKERLOTYGGAYFIN----- 1319
Qy 1007 VVELVSTALDETIDLLPTLSEGLPIIATIDGVSIGAAIKELSETNDPLLRQEI----- 1060
Db 1320 -----ESKIDVLLALSRRKAQI--SLIDAHKALTGFSFSEASLALLTGRMPGTSRV 1369
Qy 1061 ---EAKIGIMAVNLTAATAIVT---SALGTASGFSILLVPLAGISAGIPLSVNNELLIQ 1114
Db 1370 LSREVEFGPSAIVMEGATAIRAQSYDAVGLRKDFLL-----PHTVPSI--QSIIVEQ 1420
Qy 1115 DRATKVID-----YFGHI-----SLAETEGAFLLDDKIIMPODDLVLSEIDFNNSNITLQK 1166
Db 1421 AKYT-VLSWPEFYENHADKWNLANRFGA-----EDLSVHPQTFLYDTE-----GR 1465
Qy 1167 CEIMWRAEGSGG---HTLTDDIDHFFSFPSITYRKFWLSIYDVNLNKKKIDFSDKDLMWL 1222

Db 1466 C-----MGLALLYMLADD-----SVSYRL-----LQQLMFL 1492

Qy 1223 PNAPNRVFGYEMGWTPGFRSLNDGKYLDRIDHDEGQFYWFYAFIADALITKLPY 1282

Db 1493 ASLFDQENRNIPLTPADQKFLNKGSLIEWLQFGNQQLQTEGFFHTLDWIPQLMKHF 1552

Qy 1283 EDTNVRINLQGNTRSFVITTEQIRKNLSYFYGSGSYSLSPYNNIDNLVEND 1342

Db 1553 ASSTV-----KSWLITTPAHSLVLSLGMNFRVTDPNYHTDPP-SLEAAITFLER- 1602

Qy 1343 TWVIDVDNV-----KNIT-----IESDEIQK-----GELIENLSKL 1375

Db 1603 --MYQVSPAVLERVGFDEKSVTSQKLVHSLSELSQNAVFASSDLGFTSRFTTTLEEM 1660

Qy 1376 NIEDNKILANHT--INFY-----GDINESNRFISLTFISILEDI--NI 1414

Db 1661 TVRGPITMNOQRHTWATLYKIGTVQGRIDSTRRESDLN-----FLKINGDILEEFLTRT 1716

Qy 1415 IIEIDLVSKYKILLSGNC--MKLIENSSDIQKIDHIGFNGEHQKYPYSVIDNETKY 1471

Db 1717 VLDSLVDELIOQLLTKHGLEFGTTLISPSIVETAIDHVS-----LQAVKTKSR 1767

Qy 1472 NGFTDYSKGEGLTAEPFSNESIIRNIYMPDSNNLFIYSSKDLDIRINKGDKVLIGNY 1531

Db 1768 MHTIQLSUGERIF-----KLFKNSGVQSDKI-----SIDRVQLVDESDSATIDFTV 1814

Qy 1532 PKD---DMKVSLSFTIED-----TNTIKLNGVYLDENGVAQILKPM-- 1569

Db 1815 IKDQRSQKSIITVGIESLAGSPKFSASMEHVIGTGVLDLD-LGMTWVSLVQVRLVEA 1873

Qy 1570 NNAKSALNTSNLMNF---LESINIKNI 1594

Db 1874 QGKDALAVANLVNKLIALE-VSIGNV 1900

RESULT 14

Ti8501

hypotheical protein C0760c - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C:Accession: Ti8501

R:Lawson, D.; Bowman, S.; Barrell, B.

submitted to the EMBL Data Library, August 1997

A:Reference number: 218935

A:Accession: Ti8501

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3394 <LAW>

A:Cross-references: EMBL:298551; NID:e1331903; PID:e1331922; PIDN:CAB1140.1

C:Genetics:

A:Map position: 3

A:Note: C0760c

Query Match 4.4%; Score 382.5; DB 2; Length 3394;

Best Local Similarity 18.9%; Pred. No. 6.4e-06;

Matches 399; Conservative 319; Mismatches 701; Indels 687; Gaps 99;

Qy 6 KALQKQYVVKPRIOEDYVAILNALESYH--NWSESSV-----VEKYLKLDINNLT- 57

Db 699 KYLEKLNLKNIYKD-----IEKKLDLHGVIKNNKEDIYILOVEKOTLIKVISSVDY 754

Qy 58 -----NYL-----NTYKSGRNKALKKFKELYLTWEVLKNNSLTPVEKILHFTWIGQI 107

Db 755 TKMSENEHIFQNTTWNKMLNVHSSKDYNNQNNQNNQNNQNNQNNQNNQNNQNNQNI 814

Qy 108 NDTAINYNQWKDVNSD-----YTVK-----VF 130

Db 815 ENNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNNQNI 874

Qy 131 YDSNAFLINTLKKTI-----VESATNTNTLESFRENLDPEFYKFRGMELIYDKQ 183

Db 875 KDEKKTSNIIKESISLKKKHTKSIINNNDN--NNEDDDDDNDMLSVMSNDDVKNKR 932

Qy 184 KHFIDYYSQIENPEFIIDNIITKTYLSNEYSKDLBALNKYIEESLNKITPANNGNDIRNL 243

Db 933 K-----ENNKEILEEHVSFSPSN-----NEYIAHSFNSILLQLSNVIFNI 973

Qy 244 EKFADEDLRLYNOELVERWNLAAASILLISM-----LKEDGGVYLDVILPGIQDPLF 298

Db 974 E-----CKQWEYFNGSNLLSYVDYTTITIELFYRLKKYNNIFSIQIILGTQYPSIL 1024

Qy 299 KSINKPDSITNTSWEMIKLEALIMKYEVIPIGYTSKDFMDLDEEVQSFESALSSKSD--- 355

Db 1025 QKLDHGYICLDKNNKKNQNDGNSQEDDNGKKNNDG-DNKSQEDDDGNGKKNNDGDN 1083

Qy 356 KSEIFLPLDDIKVSPLEVKIAFANNVINQALISLSDSYCSDLVINOIKRYKILNDNLN 415

Db 1084 KSQ-----EDD-----YGKKKNND--DDDDSYKIELIVDELNCKKXNYTDEEL 1126

Qy 416 PSINEGTFNTMK---IFSOKLASIENEDNMFMKITN-----YLKV-----GF 458

Db 1127 YELMKGSDFDIKFKYQNPYLNHFINNINNSTIISFIHNIEDKYKYYERYFNLPMNF 1186

Qy 459 APDVRSTINLSPGV-----YTGAQDLL-----MFKDNSTN-----JHLEPE 497

Db 1187 -----SNVELSFDLLIRRFDKILRLTKYEOLESENEKIKRNEKEEYLHACIKELEMN 1241

Qy 498 LNPFPPTKISQLTEQETSLWGFNOAR--AKSQFBEYK-KGYFEGALGEDDNLD----- 550

Db 1242 LERYNEKIIDEQINEXEKINININEKYLILEKEYEYQNKNFINA-QIENLEKEKK 1299

Qy 551 FAQNTVLDRKYVSKKI-----LSSMKTRNEY-HYIVQLQCDKISYBASCNLFSPKDPYS 604

Db 1300 QLOEBEIIQDMINVKLNKNCDDIIKIYEKEQYLTLLQENKDSHNY-----LKDKFE 1352

Qy 605 SILYQKNTEGETAYYYVADAABEIKDYRIPQISNKRNIKLTFIGHSGSEFNTDFA 664

Db 1353 NLL--NL--NEKLYDH--DISLNKINTLWEEKENKKN---TF-----HMNLRAVE 1396

Qy 665 NLDVDSLSEIETIINLAKADISPKYIEINLLGCMFSPYSISAETYPKGL----- 715

Db 1397 NNLLKWKELQNKYNIILKELNERIKOINFRNNVSLSDRNRSTGSHQINNMWYN 1456

Qy 716 -----LLKIKDRVSEL-----MPSISQDSITVSANOYEVARINEE----- 749

Db 1457 NTHLGFMGASKINNINISLYSNMTHMSHRGSIKKNKEDAGNSTQARMNKNKSDTNIIN 1516

Qy 750 -----GKREILDHSGKWINKEE 766

Db 1517 NIHTNDINNMANNINNTLINSINHLYPFPHNNVNSPKMGVCDVTLAG--VNRKD 1574

Qy 767 SIIDKISSKEYISPNKKNKIIVKSKYLHELSTLLOEIRNNANSSDIDL--EKKVMLTEC 824

Db 1575 DFLNLEENEENSF--LEYEIRKS-----LQEBELCKBSEILKIKGEKNILITCI 1623

Qy 825 EI-----NVASNIDR--QIVGRIEEAKVL-----TSDSINYI----- 855

Db 1624 ETWKCFCKKGEISRLKEICKQEKKEKBFLLINKSNEDKLYINSLLCDEKDYDIVV 1683

Qy 856 -----KNPF-KLJESISDSLYD---LKHQGLDDSHFISFEDISKTENGPRIRFINKE 904

Db 1684 KOIKNNWRNEIDKLANDINEKSYEIKLKHEN---NNLINEMNLIK-----NKE 1729

Qy 905 TGNISPIETEKIFSEYATHISKEISNIKDTIFDNNVNGKLVKKNLDAHAHVNTLNSAFP 964

Db 1730 TENMTKQEE---DYIKLIKDKDTNIQNEYND-----LLEKYN-----EYVVKNNMLY 1775

Qy 965 IQSLIENYTKBSL-----SNLSVAMKVQVYA-QLFSTGLNTITDASKVVVELYSTALDETI 1019

Db 1776 NDANVLLEKHEGIEFLKENIKILOKNTYLNDFMNFQNNYVDN-----NLLKRLDQLF 1830

Qy 1020 DLLPTLSEGLPIIATIIDGVSLGAALKELSETNDPLLRQ-----EIEAKIGIMAVNLTA 1073

Db 1831 NINQDLQKH-----DINQKHEQLKYDYIEIKERLKIETKINK 1870

Qy 1074 ASTAIVTSALGIASGFSILLVFLAGISAGIPLSVNNELIQDKATKVIDYFKHISLAETE 1133

Db 927 EKDER-----DKYLEASINDYMDDKKKXYDSIESLRGSDKI---KNDQ 969
Qy 869 LYDLKHQGL-----DDSHFISPEDISKTEGFAIRFEINKETGNSIFETEKEIFSE 920
Db 970 IYQGHSSLLYYDNDNDNNNNYSSSSNNHYYILTNDKRLANDNFINNLEINNS 1029
Qy 921 YATHISKE---ISNIKDT-----IPDNVNGKLVKKNLDAHEVNTLASAFFIQSLIEY 971
Db 1030 QNKVIEKNLEVINNVKLTKTNYEQSNNTNSKDEHNISDSKSKEDTLNLS--RKSSVEY 1087
Qy 972 NT-TKESLSNLVAMKVQVVAQLPSTGLNTITDASKVVELVSTALDETIDLLPTLSEGLP 1030
Db 1088 NKKILOSTSNKSLNGAYE--NNLFS-----GKKKXKXGTVLXD-IEHINDIQDKYP 1135
Qy 1031 IIAITIDGVSLGAALKESETNDP--LLRQBIERAKIGIMAVNLTAASTAIVTSALGIASGF 1089
Db 1136 -----EDLNINCUNKVVIENEKHLPLBLE-----YNLVSSDEKF----- 1171
Qy 1090 SILLVPLAGISAGIPSLVNNELILODKATKVIDYFRHISLAETEGAFLLDDK---IIM 1145
Db 1172 -----GLNKIKDNNII-----YMKH-----QNVHNLVYDDNQKKHILF 1204
Qy 1146 PQDDLVLSEIDFNPNNSITLGCETWRAEGSGCHTLTDDIDHFFSSPSIYRKPKWLSIYDV 1205
Db 1205 DTNKNV--SIQRNNNINSVIKTHNYEVEKXNKDQORYD-----NFTCDKKKKIYNI 1254
Qy 1206 LNIKKKIDFSKDLVLPNAPRVFGYEMGT-----PGRSLONDGPK-LLDRIROHYEG 1260
Db 1255 IN--SDKDIYHNNIITYTKEKEGIGNHLNRNDKDI TNFELLKLDGVKEFLDTFKDSY-- 1310
Qy 1261 QFYWRYPAFIADALITLKPREDTNVRINDGNTRSFIVPVIITEQIRKNLSYSFYVSG 1320
Db 1311 -----IDCHNKKENI----- 1320
Qy 1321 GSYSLSLSPYMNITDLNVENDTWIDVDNVVKV-----ITIESDEIQGELIENILSKLN 1376
Db 1321 -----LNMTNKN-----KEDHQIIDVADKIFNETNMTMDNKKIYDDK-----N 1359
Qy 1377 IEDNKIILNNEHTINFYDI-----NESNPFISLTFSILEDINIIIEIDLVSXSKI-L 1428
Db 1360 VHEKK-CTHNDVIHNDMDILSTSINKNEENLF-----IDTYQKQNRIGD 1402
Qy 1429 LSGNCKLIENSDDIQQKIDHIGFGEHOKVIPYSYIDNETKYNG-----FIDYS 1478
Db 1403 IYMRINILQEDDDDDNNNNNNNNNNNNNNKILILFEYTKNDQMLHNNKNNLEGTFFSDFI 1462
Qy 1479 KKEGLTAEFSNES---IIRNIYMPDSNNLFYSSKOLKDIRIINKGDKVLLIGNYFKDD 1535
Db 1463 EKKNKIKIKNKESYHKIDESILSNEKN-----KVSLLINNKKDS 1504
Qy 1536 MKVSLSFTEDTNTIKLVGVLDENGVAQILKFWNNAKSALNTSLSMNFLESINIKNIF 1595
Db 1505 SSVD-----NNKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1552
Qy 1596 YNNLDPNIEFLDTNFI-----ISGNSISIQPELICDDKNIQPYFINEKIKETSYTL 1648
Db 1553 -NNNNNNDSFKONNLINNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1611
Qy 1649 YVGNFQNL-----IVEP-----SVHLDSDGNSISSTVINFSOKLYGIDR 1687
Db 1612 LINNHPNLNSMRKRTIEPLKINGKNKILKDLKIQOQVERKIRKYKIOMDQ 1664

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 16:44:46 ; Search time 56.875 Seconds
(without alignments)
7840.786 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1700
Perfect score: 8675
Sequence: 1 NMLVKAQLQKVMYKFRIQ.....VLYGIDRVYKVIAPNLYT 1700

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1069545 seqs, 262320428 residues

Total number of hits satisfying chosen parameters: 1069545
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6759.5	77.9	2366	14	US-10-011-366-10 Sequence 10, Appl
2	6759.5	77.9	2366	15	US-10-354-774-10 Sequence 10, Appl
3	6759.5	77.9	2366	15	US-10-271-012-10 Sequence 10, Appl
4	4467	51.5	2710	14	US-10-011-366-6 Sequence 6, Appl
5	4467	51.5	2710	15	US-10-354-774-6 Sequence 6, Appl
6	4467	51.5	2710	15	US-10-271-012-6 Sequence 6, Appl
7	2204	25.4	556	12	US-10-463-957-1 Sequence 11, Appl
8	2193	25.3	556	12	US-10-463-957-11 Sequence 11, Appl
9	2190	25.2	556	12	US-10-463-957-12 Sequence 12, Appl
10	2051.5	23.6	522	12	US-10-463-957-9 Sequence 9, Appl
11	1979	22.8	500	12	US-10-463-957-3 Sequence 3, Appl
12	1734	20.0	420	12	US-10-463-957-5 Sequence 5, Appl
13	769	8.9	170	12	US-10-463-957-7 Sequence 7, Appl
14	512.5	5.9	3169	14	US-10-114-170-257 Sequence 257, App
15	431.5	5.0	4688	12	US-10-282-122A-76865, A

16	347	4.0	6641	12	US-10-282-122A-70580	Sequence 70580, A
17	340.5	3.9	5005	12	US-10-282-122A-76871	Sequence 76871, A
18	336	3.9	4620	12	US-10-282-122A-68921	Sequence 68921, A
19	312	3.6	1639	14	US-10-087-464-10	Sequence 10, Appl
20	309.5	3.6	1937	15	US-10-369-493-2070	Sequence 2070, Ap
21	304	3.5	2184	14	US-10-304-095-6	Sequence 6, Appl
22	296	3.4	2017	15	US-10-452-024-146	Sequence 146, App
23	290	3.3	1182	12	US-10-282-122A-53445	Sequence 53445, A
24	289	3.3	2575	12	US-10-282-122A-72134	Sequence 72134, A
25	288	3.3	1875	15	US-10-369-493-22285	Sequence 22285, A
26	287	3.3	1610	14	US-10-155-533-9	Sequence 9, Appl
27	286.5	3.3	2046	12	US-10-282-122A-72132	Sequence 72132, A
28	285.5	3.3	1948	14	US-10-032-585-7611	Sequence 7611, Ap
29	281.5	3.2	1881	14	US-10-032-585-7646	Sequence 7646, Ap
30	281.5	3.2	1196	12	US-10-282-122A-52737	Sequence 52737, A
31	281	3.2	3692	12	US-10-282-122A-71235	Sequence 71235, A
32	279	3.2	1178	12	US-10-282-122A-52434	Sequence 52434, A
33	275.5	3.2	1279	12	US-10-282-122A-52455	Sequence 52455, A
34	274	3.2	1847	14	US-10-032-585-7538	Sequence 7538, Ap
35	273.5	3.2	1928	15	US-10-369-493-22025	Sequence 22025, A
36	271	3.1	1494	12	US-10-282-122A-47128	Sequence 47128, A
37	269.5	3.1	2167	9	US-09-801-368-56	Sequence 56, Appl
38	269.5	3.1	2167	15	US-10-369-493-1773	Sequence 1773, Ap
39	268.5	3.1	1184	12	US-10-282-122A-53254	Sequence 53254, A
40	267.5	3.1	1837	15	US-10-369-493-22734	Sequence 22734, A
41	267.5	3.1	6713	12	US-10-282-122A-43811	Sequence 43811, A
42	266.5	3.1	1465	12	US-10-282-122A-47271	Sequence 47271, A
43	264.5	3.0	3533	12	US-10-282-122A-70177	Sequence 70177, A
44	262	3.0	2013	12	US-10-282-122A-60608	Sequence 60608, A
45	259.5	3.0	1847	15	US-10-369-493-1075	Sequence 1075, Ap

ALIGNMENTS

RESULT 1

US-10-011-366-10
; Sequence 10, Application US/10011366
; Publication No. US20030054493A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Kink, John A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES.
; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
; DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/011,366
; FILING DATE: 16-No. US20030054493A1-2001
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,310
; FILING DATE: 23-OCT-1997
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 24-OCT-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPND-01121
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-011-366-10

Query Match 77.9%; Score 6759.5; DB 14; Length 2366;
Best Local Similarity 76.6%; Pred. No. 0;
Matches 1302; Conservative 191; Mismatches 206; Indels 1; Gaps 1;
QY 1 MNLVKAQLOKQVYKFRIOEDYVAILNALLEEYHNSSSVVEKYKLDINDNLDNYL 60
DB 1 MSLVNRKLEKXANVRPTQDEYVAILDALEEYHNSSSVVEKYKLDINDNLDNYI 60
QY 61 NTKKSGRNKALKKPEYLTWVLELKNLSLTPVEKNLHPTWIGQINDTAIYNQWKD 120
DB 61 DTKKSGRNKALKKPEYLTWVLELKNLSLTPVEKNLHPTWIGQINDTAIYNQWKD 120
QY 121 VNSDYTVKVFYDSNAFLNTLTKTTVESATNNTLESFRENLDPEFYNNFYKRMELIY 180
DB 121 VNSDYTVKVFYDSNAFLNTLTKTTVESATNNTLESFRENLDPEFYNNFYKRMELIY 180
QY 181 DKQHFIDYKQIBENPEFIDNIKITLSNEYSKOLEALNKYIEBSLNKITANNNDI 240
DB 181 DKQNFINYKQARENPELIDDIIVKYLSEYKEIDELNTYIEBSLNKITQNSGNDV 240
QY 241 RNLEKPAEDLVRLNQBELVERNLAAASDILRLSMKDEGCVGLVDVILPGIQDPLFKS 300
DB 241 RNFEFKNGESFNLYEQELVERNLAAASDILRLSMKDEGCVGLVDVILPGIQDPLFES 300
QY 301 INKPSITNTSHMILKLAIMKYKEYIPGYTSKQFMDLDEVSQSFESALSKSKSEIF 360
DB 301 IEKPSVTVDFWEMTKLAIMKYKEYIPGYTSKQFMDLDEVSQSFESALSKSKSEIF 360
QY 361 LPLDDIKVSPLEKIAFANNSVINQALSLKDSYCSGLVINQIKRYKILNDNLNPSINE 420
DB 361 SSLGDEWASPLEKIAFANNSVINQALSLKDSYCSGLVINQIKRYKILNDNLNPSINE 420
QY 421 GTDFNTTKIFSDKLASISNEDNMFMKITNYLVKGFAPDVRSTINLSGPGVYTGAYQD 480
DB 421 DNDFTNTTTFDISMAEANADNGRFMELGKYLAVGFFPDVKTINLSGPEAYAAAYQD 480
QY 481 LLMFQDNGSTNHLBPENRFPKTKISQLTQEBITSLWFPNQARAKSOPEEYKKGYPE 540
DB 481 LLMFQEGSMNIHLIEADLENFBISKTNSQSTEQEMASLWSPFDDARAKAOPPEYKRYNFE 540
QY 541 GALGDDNDLPAQNTVLDKDYVSKYKILSMKTRKEYTHYIYVLOGDKIYSYASCNLPSK 600
DB 541 GSLGDDNDLPAQNTVLDKDYVSKYKILSMKTRKEYTHYIYVLOGDKIYSYASCNLPSK 600
QY 601 DPVSSILYQKNEGSETAYYYVADAEIKEDIKYRIPYQISNKNIKLTFIGHGKSEPT 660
DB 601 TPYDSVLPQKNEIDSEIAYYVPGDGEIOEDIKYKIPSIIDRPKIKLTFIGHGKDEPT 660
QY 661 DTPANLDVDSLSEITETINLAKADISPKYIEINLLGNMFPYSISABETPGKLLLIK 720
DB 661 DTFAGPDVDSLSEIEAIDLAKEDISPKYIEINLLGNMFPYSINVEETPGKLLLVK 720
QY 721 DRVSELMPSISQDSITVSANQEVYRINEEKEILDHSGKWNKEESIKDISKEYISF 780
DB 721 DKISELMPSISQDSITVSANQEVYRINEEKEILDHSGKWNKEESIKDISKEYISF 780
QY 781 NPKENKIIVKSXYLHSLTLQEIERNANSDDILEKKVMTCEINVASNIDRQIVEGR 840

DB 781 NPKENKIIVKSXYLHSLTLQEIERNANSDDILEKKVMTCEINVASNIDRQIVEGR 840
QY 841 IEEAKNLTSDSINIVKNEFKLIESISDLYDLKHONGLDSDSHFISFEDISKTENGFRIRF 900
DB 841 IEEAKNLTSDSINIVKNEFKLIESISDLYDLKHONGLDSDSHFISFEDISKTENGFRIRF 900
QY 901 INKETGNSIFITEKEIFSEYATHISKEISINIKDTIFDNVNGKLVKKVNLDAAEVNTLN 960
DB 901 INKETGESIFVETEKTFSEYANHIITEISIKGTIFDTVNGKLVKKVNLDTTHEVNTLN 960
QY 961 SAFFIQSLIEYNTTKESLSNLSVAMKQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
DB 961 AAFPIQSLIEYNSKESLSNLSVAMKQVYAQLFSTGLNTITDASKVVELVSTALDETID 1020
QY 1021 LLPTLSEGLPIIATIIDGVSIGAAIKELSETNDPLLRQIEIAKIGIYAVNLTAASATV 1080
DB 1021 LLPTLSEGLPIIATIIDGVSIGAAIKELSETNDPLLRQIEIAKIGIYAVNLTAATIT 1080
QY 1081 SALGTASGFSILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETEGAFTLLD 1140
DB 1081 SGLGTASGFSILLVPLAGISAGIPSLVNNELVLRDKATKVIDYFKHISLAETEGAFTLLD 1140
QY 1141 DKIMPDQDLVSEIDFNNNSITILGKCIWRAEGSGHTLTDIDHFFSSSITYRKPLW 1200
DB 1141 DKIMPDQDLVSEIDFNNNSITILGKCIWRAEGSGHTLTDIDHFFSAPSITYREPHL 1200
QY 1201 SIYDVNLTKKEIKIDFSKOLMVLNPNAPRVFGVEMGTGPFSLDNDGKLDRIEDHYEG 1260
DB 1201 SIYDVLEQKEBLDSKOLMVLNPNAPRVFAWETGTPGLSLNDNDGKLDRIEDHYEG 1260
QY 1261 QFWRYFAFIADALITLKPVEDTNVRLINDGNTRSFIVPVTTEIQIKNLSYFYSG 1320
DB 1261 EFWRYFAFIADALITLKPVEDTNVRLINDGNTRSFIVPVTTEIQIKNLSYFYSG 1320
QY 1321 GVSLSLSYPNNKIDILNVENDTWIVDNNVKNITIESDETOQKGLIENILSKLNIEDN 1380
DB 1321 GYVALSLSQYNNWGINIELSESDVWIIDVNVVRDVTIESDKIKKGLIEGILSTLSIEN 1380
QY 1381 KILANNHTNFYGDINESNRFSLTFSILEIDNIIIEIDLVSYSKILLSGCMKLIENS 1440
DB 1381 KIILNSHEINFSGEVNGSGFVSLTFSILEGINAIEVDLLSKYKLLISGELKILMLNS 1440
QY 1441 SDIQOIKDIHGNGEHOKYIPYSYIDNETKYNFIDYSKKEGLFTAEFNSESIIRNIYMP 1500
DB 1441 NHIQOIKDIYIGVSELSQKNIPIYSFVDSSEKNGFNGSTKEGLFVSELDDVVLISKVMD 1500
QY 1501 DSNNLFIYSKDKIDIRIINKGDKVLLIGNYFKDMKVSLSFTIEDTNTIKLNGVYLDEN 1560
DB 1501 DSKPSFGYISNNLKDVKVITKONVILTGYYLKDDIKISLTLQDEKTIKLSVHLDSS 1560
QY 1561 GVAQILKFMNNAKSALNTSNLMNFIENIKNIFVNNLDPNIEFLDTPNFIISGNSIG 1620
DB 1561 GVAELIKFMNR-KGNTNTSDLSMFLSNMKSIFVNFQSNIKFILDANFIISGTTSIG 1619
QY 1621 QFELICDKKNQOPYINFKIKETSYTLVGNRONLIVPSVHLDSDSGNISSTVINFSOK 1680
DB 1620 QFELICDENDNIOPYIKETLETNTYLVGNRONMIVEPNYDLDSDSGDISSTVINFSOK 1679
QY 1681 YLYGIDRYNKKVIAANLNT 1700
DB 1680 YLYGIDSCVKNKVIISPNIT 1699

RESULT 2
US-10-354-774-10
; Sequence 10, Application US/10354774
; Publication No. US20030215468A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; ; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; ; Botulinum Neurotoxin

NUMBER OF SEQUENCES: 82
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Medlen & Carroll
 STREET: 220 Montgomery Street, Suite 2200
 CITY: San Francisco
 STATE: California
 COUNTRY: United States of America
 ZIP: 94104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/354,774
 FILING DATE: 30-Jan-2003
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/704,159
 FILING DATE: 28-AUG-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Ingolia, Diane E.
 REGISTRATION NUMBER: 40,027
 REFERENCE/DOCKET NUMBER: OPD-02304
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 705-8410
 TELEFAX: (415) 397-8338
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2366 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 10:
 US-10-354-774-10

Query Match 77.9%; Score 6759.5; DB 15; Length 2366;

Best Local Similarity 76.6%; Pred. No. 0;
Matches 1302; Conservative 191; Mismatches 206; Indels 1; Gaps 1;

QY 1 MNLVKAQLOQWVYKFRIGQDEVAALNAAEEVHNSSESVVEKYLKLDINLTNDVL 60
 DB 1 MSLVNRKQLEKAAVVRPTQDEVAALDAAEEVHNSSESVVEKYLKLDINLTNDVL 60
 QY 61 NTYKSGRNKALKKFKKYLTVMEVLELKNLSLTPVEKNLHPITWIGGQINDTAINYNQWKD 120
 DB 61 DTYKSGRNKALKKFKKYLTVMEVLELKNLSLTPVEKNLHPITWIGGQINDTAINYNQWKD 120
 QY 121 VNSDYTVKVDNAPLINTLTKTIVESATNWTLESFRENLDNPEFDYKFKRMEIYY 180
 DB 121 VNSDYTVKVDNAPLINTLTKTIVESATNWTLESFRENLDNPEFDYKFKRMEIYY 180
 QY 181 DKQKHFIDYKSOJTEENPERFIIDNIKTLYLSNEYSKDLALNKYIEESLNKITAANGNDI 240
 DB 181 DKQKFNINYKAQREAPELIIDIVKTYLSNEYSKEIDELNTYIEESLNKITAANGNDV 240
 QY 241 RNLEKFADEDLRLYNQELVERMNLAAASDILRISMLKEDGGVYLDVILPGIQDLPFKS 300
 DB 241 RNLEKFADEDLRLYNQELVERMNLAAASDILRISMLKEDGGVYLDVILPGIQDLPFKS 300
 QY 301 INKPDSTNTWEMIKELAIMKYEYIPGYTSKNFMDLDEVOVSFESALSKSKSEIF 360
 DB 301 IEKPSVTVDFWEMTKELAIMKYEYIPEYTSSEHFDMLDEVOVSFESALSKSKSEIF 360
 QY 361 LPLDDIKVPLEVIAFANNVINOALISLKDSVCSPLVINQIKNRYKIINDNLNPSINE 420
 DB 361 SSLGDMASPLEVIAFANNVINOALISLKDSVCSPLVINQIKNRYKIINDNLNPSINE 420
 QY 421 GTDFNTTKIFSDKLASINEDNMFMKIKNYLVKGFAPDVRSTINLSPGYITGAYOD 480
 DB 421 DNDFTNTNTFIDSIMAFANADNCRFMWELGKYLKLVGFFEDVKTINLSPGEYAYAYOD 480
 QY 481 LLMFKDNSTNTHLEPELRNPEFPKTKISQLTQESITSLMSFNQARAKQFEBYKKGYPE 540

DB 481 LLMFKEGSMNIHLIEADLRNFEISKNTISQTEQEMASLWSFDDARAKAQFYEYKRNYPE 540
 QY 541 GALGEDDNLDPACNTVLDKDYVSKYLSSMKTENKEYIHVIYVLOQDKIYSBASCNLPSK 600
 DB 541 GSLEDNDLDPSONIVVDKEYLLEKISSLARSGERYIHVIYVLOQDKIYSBAACNLPAK 600
 QY 601 DPYSSILYQKNIEGSETAYYYYVADAEIKEDIKYRIPYQISNKRNIKLTFFIGHKSEFNT 660
 DB 601 TPYDVLVFNQVIEDSEIAYVYVPGDGEIQEIDKVKIPSIISDRPKIKLTFIGHGKDEPNT 660
 QY 661 DTPANLVDLSSEIEITILNLAADISPKYIEINLLGCMNFSYSISAEITYGKLLKLIK 720
 DB 661 DIFAGFDVDSLSTEIEAADIADLAKEDISPKSIEINLLGCMNFSYSINVEITYGKLLKLVK 720
 QY 721 DRVSELMPSISODSITVSANQYEVNRNEEGKEIHDHSGWKINKEESIKDSSKEYISF 780
 DB 721 DKISELMPSISQDSIIVSANQYEVNRINSEGRHELLDHSGEWINKKEESIKDISKEYISF 780
 QY 781 NPXENKIIVKSKYLHELSTLLQEIARNANSSDIDLEKVMLTECEINVASNIDROIVEGR 840
 DB 781 NPXENKITVSKNLPSELSTLLQEIARNANSSDIDLEKVMLTECEINVASNIDTQIVEER 840
 QY 841 IEEAKNLTSDSINYIKNEFKLIESISDSLYDLKHQNGLDSDSHFISPEDISKTENGEPFRP 900
 DB 841 IEEAKNLTSDSINYIKDEPKLIESISDALCDLKQQNELEDSDHFISPEDISETDEGFSIRF 900
 QY 901 INKETGSIETETEKEIFSEYATHISKEISNTKDTIFDNVNGKLVKKNLDAAEVNTLN 960
 DB 901 INKETGESIFVEYTEKIFSEYANHTIEESKIGTIFDVNGKLVKKNLDTTHEVNTLN 960
 QY 961 SAPFIQSLIEYNTTKSLSNLSVAMQVYVYQALFSTGLNTITDAGKVELVSTALDETID 1020
 DB 961 AAFPIQSLIEYNSKESLSNLSVAMQVYVYQALFSTGLNTITDAGKVELVSTALDETID 1020
 QY 1021 LLTSLSEGLPIIATIIDGVSGLGNAIKELSTNDPLLRQIEAKIGTMAVNLTAATAIYT 1080
 DB 1021 LLTSLSEGLPIIATIIDGVSGLGNAIKELSTNDPLLRQIEAKIGTMAVNLTAATAIYT 1080
 QY 1081 SALGIASGPIILLVPLAGISAGIPSVNNELILODKATKVIDYFKHISLAETEGTAFTLD 1140
 DB 1081 SSGIASGPIILLVPLAGISAGIPSVNNELVLRDKATKVDFYFKHISLAETEGTAFTLD 1140
 QY 1141 DKIMPDDLVLSIEDFNNNNSITLKGCEIWRAGGGSGHTLTDIDHFFSPSITYRKPWL 1200
 DB 1141 DKIMPDDLVLSIEDFNNNNSITLKGCEIWRAGGGSGHTLTDIDHFFSPSITYRKPWL 1200
 QY 1201 SIYDVNLKKEKIDFSKDLMLPNAPRVGYEMGWTGFRSLDNDGTCLLDRIEDHVEG 1260
 DB 1201 SIYDVNLKKEKIDFSKDLMLPNAPRVGYEMGWTGFRSLDNDGTCLLDRIEDHVEG 1260
 QY 1261 QFYWRYFAFTADALITKLPREVDTNVRINLGNTPSPVPTTBOIEKNLSYFYGSG 1320
 DB 1261 EFWRYFAFTADALITKLPREVDTNVRINLGNTPSPVPTTBOIEKNLSYFYGSG 1320
 QY 1321 GSYLSLSPYNNMIDNLVENDTWIVDNNVKNITIESDEIQKGLIEHLNLSKLIEN 1380
 DB 1321 GTVALSLSQNVGINTSESDWIIDVNNVDRVTIESDKIKGDLIEHLNLSKLIEN 1380
 QY 1381 KIILNHTINPYGDINESRPFSLTSLIEDINIIIEIDIVSKSYKILLSGCMKLIENS 1440
 DB 1381 KIILNHTINPYGDINESRPFSLTSLIEDINIIIEIDIVSKSYKILLSGCMKLIENS 1440
 QY 1441 SDIOQKIDHIGFNGEHQKIPYSDINETKNGFIDYSKEGLTAEFENESIIRNIYMP 1500
 DB 1441 NHIOQKIDHIGFNGEHQKIPYSDINETKNGFIDYSKEGLTAEFENESIIRNIYMP 1500
 QY 1501 DSNLFTYSSKDLKDIRIINKGDVLLIGNYFKODMKVSLSTIEDTNTIKLVGVLBN 1560
 DB 1501 DSKPSPFYGSNNLKVKITKDNVNLITGYLKKDIIKISLSLTLODEKTIKLSVHLDES 1560
 QY 1561 GVAQILKFMNNAKSALNTSNLNMNFIENIKNIFANNLDPNTEFLDINFIISGNSIG 1620

Db 1561 GVAEILKPMNR-KGNTWNTSDLSMFLSMNIXISFVNFLQSNKIFILDFANFIISGTTISG 1619

Qy 1621 QFELICDKDNIOPIYFINKETSYLYVGNQNIIVFSPYHLDGSGNISSTVINFSOK 1680

Db 1620 QFELICDENIOPIYKFTNLTNTYLYVGNQNIIVFSPYHLDGSGNISSTVINFSOK 1679

Qy 1681 YLYGIDRYVKNVLIAPNLYT 1700

Db 1680 YLYGIDSCVKNWISPNYIT 1699

RESULT 3

US-10-271-012-10

Sequence 10, Application US/10271012

Publication No. US20030219457A1

GENERAL INFORMATION:

APPLICANT: Williams, James A.

Thalley, Bruce S.

TITLE OF INVENTION: Multivalent Vaccine For Clostridium Botulinum Neurotoxin

NUMBER OF SEQUENCES: 82

CORRESPONDENCE ADDRESS:

ADDRESSEE: Medlen & Carroll

STREET: 220 Montgomery Street, Suite 2200

City: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/271,012

FILING DATE: 15-Oct-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/704,159

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Ingolia, Diane E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPHD-02304

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 2366 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 10:

US-10-271-012-10

Query Match 77.9%; Score 6759.5; DB 15; Length 2366;

Best Local Similarity 76.6%; Pred. No. 0;

Matches 1302; Conservative 191; Mismatches 206; Indels 1; Gaps 1;

Qy 1 MSLVNRKQLEKMANFRQDEYVAILDALEBYHNMSENTEVVEKYLKLDINSITDIYI 60

Db 1 MSLVNRKQLEKMANFRQDEYVAILDALEBYHNMSENTEVVEKYLKLDINSITDIYI 60

Qy 61 NTYKSGRNKALKKFKLEYLTMEVLELKNNSLTPEKXNLHFWIGQINDTAINYNQWKD 120

Db 61 DTYKSGRNKALKKFKLEYLTMEVLELKNNSLTPEKXNLHFWIGQINDTAINYNQWKD 120

Qy 121 VNSDYTVKYFDNSAFINLTNLTKTIVESATNTLTESFRENLDNDFDYNKFKRMEIY 180

Db 121 VNSDYVNVFYDNSAFINLTNLTKTIVESATNTLTESFRENLDNDFDYNKFKRMEIY 180

Qy 181 DKQKHFDIYKSGIENPEFIIDNIITKYLNSKDLALNKYIEESLNKNTANNNDI 240

Db 181 DKQXNFYNYKAQREENPELIIDIVKTYLSNEYSKEIDELNTYIEESLNKNTQNSGNDV 240

Qy 241 RNLKFADEDLVRLYNOELVERWNLAASDILRISMLKEDGGVVDVLDLPGQPDLFKS 300

Db 241 RNFPEFKNGESFNLYQEDELVERWNLAASDILRISMLKEDGGVVDVLDLPGQPDLFKS 300

Qy 301 INKPDSTINTSWEMIKLEAMKWEYIPGYTSKNFMDLDEEVORFSPFSAKSDKSEIF 360

Db 301 IEKPSVTVDFWENTKLEALWKYKEYPEYTSBHFMDLDEEVQSSPFSAKSDKSEIF 360

Qy 361 LPLDDIKVSPLEVKIAFANNVINOALISLKSVCSDLVINOIKRYKILNDNLNPSINE 420

Db 361 SSLGMEASPLEVKIAFNKSGIINOGLISVKDSYCSNLIVKQIENRYKILNNSNPAISE 420

Qy 421 GTDFNTWKIFSDKLASISNEDNMFMKITYNYLKVGFADPDRVSTINLSGPGVVTGAYOD 480

Db 421 DNDFTNTTNTFIDSIMAEANADNGRFVWELGKYLVRGVFFPDVKTTINLSGPEAYAAAYOD 480

Qy 481 LLMFKDNTNTHLPELPELPEFKTKISOLTEQIETSLWSFNQARAKSQPEEYKGYFE 540

Db 481 LLMFKEGSMNIHLHEADLRNFEISKTNISQTEGEMASLWSPDDARAKAQPEEYKRYFE 540

Qy 541 GALGEDDNLDPQNTVLDKDYVSKKILSSMKTRKEYTHYIVLQGDKISYEASCNLPSK 600

Db 541 GSLGEDDNLDPSONIVVDKEYLLEKISLARSSEGYTHYIVLQGDKISYEAACNLFAK 600

Qy 601 DPYSILYQKNIESETAYYYYVADAETKEIDKVRIPYQISNKNKIKLTFHGHKSEPT 660

Db 601 TPYDSVLFQKIEDSEIAYYYNPGDGEIQEIDKVKIPSIISDRPKIKLTFHGHKDEFT 660

Qy 661 DTFANLVDVSLSSBIETILNLAADISPKYIEINLLGCMFSPYSISABETYPGKLLKIK 720

Db 661 DIFAGFDVDSLSTEIEAADIKEDISPKSIEINLLGCMFSPYSINVEETYPGKLLKLVK 720

Qy 721 DRVSELMPSISODSITVGSANQVEYVINEEGREIILDHSGKWNKEESIIDISKEYISF 780

Db 721 DKISELMPSISQDSIIVSANQVEYVINEEGREILDHSGEWNKEESIIDISKEYISF 780

Qy 781 NPENKIIIVKSKYLHELSTLQEIERNANSDDIDLEKKVMLTECEINVASNIDROIVEGR 840

Db 781 NPENKIIIVKSKXNLPESLTLQEIERNANSDDIDLEKKVMLTECEINVASNIDTQIVEER 840

Qy 841 IEAKNLTSDSINYINKNEPKLIESDSLYLKHQGLDDSHFISFEDISKTENGFRFRF 900

Db 841 IEAKNLTSDSINYINKNEPKLIESDLCQKQNELEDSHFISFEDISETDFGSIRF 900

Qy 901 INKETGNSIFITEKEIPEYATHISKETISNIDTFDMVNGKLVKKNLDAAEHVNTLN 960

Db 901 INKETGESIFVETEKTFIFSEYANHTIEESKIKGIFDTVNGKLVKKNLDTTHEVNTLN 960

Qy 961 SAPFIQSLIEYNTTKESLSNLSVAMKVQVYAOQFSTGLNTITDASKVVELVSTALDETID 1020

Db 961 AAFPIQSLIEYNSKESLSNLSVAMKVQVYAOQFSTGLNTITDAKKVVELVSTALDETID 1020

Qy 1021 LPLTSEGLPIIATIIDGVSLGAAIKELSETNDPLLRQIEAKIGIMAVNLTAATAIYT 1080

Db 1021 LPLTSEGLPIIATIIDGVSLGAAIKELSETNDPLLRQIEAKIGIMAVNLTAATAIYT 1080

Qy 1081 SALGIASGSSILLVPLAGISAGIPSVNNELLLODKATVIDYFKHISLAETEGATFLD 1140

Db 1081 SSLGIASGSSILLVPLAGISAGIPSVNNELLVLRDKATVKDYFKHVSJVEVEGVTFLD 1140

Qy 1141 DKTIIMQDDVLVLEIDDFNNNSITFLGCEIWRABGGSGHTLTDIDHFFSSPSITYRKPLW 1200

Db 1141 DKTIIMQDDVLVLEIDDFNNNSIVLGCEIWRMEGSGHVTDDIDHFFSAPSTIYREPHL 1200

Qy 1201 SIYDVNLNKKKEKIDFSKDLMLVFNAPNRFVGYEMGWTGFRSLDNDGTKLLDIRHYEG 1260

Db 1201 SIYDVLEQKEELDLSKDLMLVFNAPNRFVGTGTGFRSLDNDGTKLLDIRNYEG 1260

Qy 1261 QFYWRYPFAFTADALITKLPKRYEDTNRINLDGNTSFIVPVITTTQIRKNLSYFVSGS 1320

```

1261 EFYWRVFAIADALITLTKPRVEDTNRINLDSNTRSFIVPIITTEVIREKLSYFYGSG 1320
1321 GYSLSLSPNMWIDNLVENDTWIDVDNVAKNITIESEDOXGELIENILSKLINIEN 1380
1321 GTVALSLQYMGINIELSESDVMIIDVDNVEDVTIESDKIKGDLIEGILST:SIEN 1380
1381 KIILNHTINFYGDINESNRFISLTSEILEDINIIIEIDLVSXKILLSGCMKLIENS 1440
1381 KIILNHEINFSEVNGSFGVSLTSEIIEGINAIEVLLSXSXYLLISGBELKILMLNS 1440
1441 SDIQKIDHTIGFNGEHQKIPIYSYIDNETKYNQFIDYSKKEGLFTAEFSESIRNIYMP 1500
1441 NHIOQKIDYIGFNSQLKNIPYFVDSGKENGFGINGSTKEGLFVSELDDVVLISKVYMD 1500
1501 DSNNLFIYSKDKIDRIINKGVKLLIKNFYKODMKVLSPTIETNTIKLVNGVLDEN 1560
1501 DSKPSFGYGNLKDVKVITKDNVILITGYLLKDDIKISLSLTLODEKTIKLSNVELDS 1560
1561 GVAQILKPMNKAASLNTSLSNLELESINIKIFYNLDPNLEFILDNFIIISGNSIG 1620
1561 GVAEILKPMNR-KGNTNTSLSLMSFLESMMIKSIFVNFLOSNIKFILDANFIISGTSIG 1619
1621 QFELICDKXNIOPYFINFKIKETSIVLYVGNRQNLIVEPSYHLDSDSGNISSTVINFSOK 1680
1620 QFEEICDENDNIQPIFYKFNTELTETNYLYVGNRQNLIVEPNYLDSDSGNISSTVINFSOK 1679
1681 YLYGIDRYNKKVIAAPNLYT 1700
1680 YLYGIDSCVNKWISPIYNT 1699

RESULT 4
US-10-011-366-6
; Sequence 6, Application US/10011366
; Publication No. US20030054493A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Kink, John A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
; DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESS: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/011.366
; FILING DATE: 16-NO. US20030054493A1-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/957.310
; FILING DATE: 23-OCT-1997
; APPLICATION NUMBER: US 08/329.154
; FILING DATE: 24-OCT-1994
; APPLICATION NUMBER: US 08/161.907
; FILING DATE: 02-DEC-1993
; APPLICATION NUMBER: US 07/985.321
; FILING DATE: 04-DEC-1992
; APPLICATION NUMBER: US 07/429.791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01121

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-011-366-6

Query Match      51.5%; Score 4467; DB 14; Length 2710;
Best Local Similarity 50.7%; Pred. No. 9.8e-271;
Matches 871; Conservative 366; Mismatches 440; Indels 42; Gaps 15;

QY 1 MNLVKAQQLQWVYVVKRIQDEYVAINALDEEYHNMSESSVVEKYLKLDINNLTNYL 60
DB 1 MSLISKEELIKLAY-SIRPRENEYKTLTNDLEYNKLTNNENKYLQKLKLNESIDVFM 59
QY 61 NTRYKSGRNKALKKFKKYLTVLEVLKQNSLTVPVEKNLHFIWIGQINDTAINIOWKD 120
DB 60 NKYTSSRNRLSNLKKDILKEVILIKNSNTSPVEKNLHFVWIGREVSIDALEYIKQWAD 119
QY 121 VNSDYTVKVFYDSNAFLINTLKTIVESATNTLESFRENLDPEFDYKFKRMEIY 180
DB 120 INAEYNIKLYWYDEAFVLTLLKAVESSTTEAQLLEBEIQNFQDNMFKYKRMFEIY 179
QY 181 DKQHFIDYKSOJEENPEFTIDNIITKYLSEYNSKOLEALNKYIEESLNKITANNNDI 240
DB 180 DRQRFINYKSNQINKPTVPTIDDIKSHLVSEYNRDETVELSYRTNSLAKINSNHGIDI 239
QY 241 RNLEKFADEDLVRLYNQELVERWNLAASDILRISMLKEDGGVYLDVILPGIQDLPFKS 300
DB 240 RANSLFTEQELNIVSQELLNRGLAAASDIVRLALKNGVYLDVDMPLGHSIDLPKT 299
QY 301 INKPSDITNTSWEMIKLEAIKMYKEYIPGYSKQKPMLEDEEVQSPSALSSKSDKSEIF 360
DB 300 ISRPSSIGLDRWEMIKLEAIKMYKYNNTYSENFDKLQQLKONFKLIIESKSEKSEIF 359
QY 361 LPLDDIKVSPLEVKIAPANNVINOALISLKDSCDLVINOIKRYKILNDNLNPSINE 420
DB 360 SKLENLVSDLEIKIAPALGVSINOALISKOGSLTNTLVTEQVKNRQFLNQHNLPALES 419
QY 421 GTDFNTTMMKPSDKLASISNEDNMFMKITYNLKVGFPADPVRSTINLSGPGVVTGAYOD 480
DB 420 DNNFTDTTKIPIHDSLPNSATAENSFLLTKIAPYLVQVGFMPARSTISLSGPGASAYD 479
QY 481 LLMFKNSTNIHLEPELNFPEPKTKISQLTQEITSLMSFNQARAKSQPEEYKGYFE 540
DB 480 FINLQNTIETKTLKASDLIEFKFPENNLSQLTQEINSLSFQASAKYQPEKYVRDYG 539
QY 541 GALGEDDNLDPQNTVLDKDY-VSKKILSS--MKTRNKEYIHYIVLOQDKISYEASCNL 597
DB 540 GSLSEDNGVDENKNTALDKNLLNKKIPSNNVESAGSNVYHYIQLGGDDISYEATCNL 599
QY 598 FSKOPYSIILYKQNIEGSETAYVYVAD--AEIKEDKYRIPYQISNKNIKLFIHGKG 655
DB 600 FSKNPKNSIIIRQNM--NESAKSYFLSDGSEIIELNKIRIPERLKNKVKYVFIHGKG 657
QY 656 SEFTDTPTANLDDVDSLSEIETILNLAKADISPKYIEINLILGCMNFSYSIAEETYPGL 715
DB 658 DEFTSEFARLSVDSLSEISFLDTIKDISPKNVEVNLGCMNFSYDFNVEETYPGL 717
QY 716 LKIKORVSELMPSISQSDITVSANQVEVRINEGKREILDHSGKWINKESESIDKSSK 775
DB 718 LLSIMDKITSTLPVKNKSITIGANQVEVRINSEGRKELLASHGKWINKESEAIMSDLSK 777
QY 776 EYISFNPKENKLIYKSKYLHELSTILLQIRNANSSDIDLEKVMLECEINVASNDRQ 835
DB 778 EYIFPDSIDNKLKAKSKNIPGLASISEDIKTLILDASVSPDTKFIPLNKLNISSIGDY 837
QY 836 IVEGRIEBAKNLTSDSINYIKNEFKLIESISDSLYDLKHQGLDDSHFISFEDISKTEG 895

```

838	IYYEKLEPYKNI	IHNSIDBLIDFENLLENVSDLEYELKKNLNDKYLISFEDISQKNS	897	
896	FRIRFINKETGNS	FIFETEKEIFSEYATHSIKESINI	KOTIFDNVNGKLVKKVNLDAAE	955
898	YSRVFINKGNSV	YVETEKEIFSKYSEHITKESITKNSIT	ITDVGNNLDMNIQLDHTSQ	957
956	VNTLSAFTIQS	LIEYNTTKESLNSIVAMKVQVYAQLFSTGLNTITDASKVVELVSTAL	1015	
958	VNTLNAAFIQS	LIDYSSKNKVDLNDLSTSVKQVYAQLFSTGLNTIYDISQVLNISNAV	1017	
1016	DETIDLLPSL	SEGLPIATIIDGVSGAAIKELSETNDPLLRQETBAK	GIMAVNLTAAS	1075
1018	NDTINVLPIT	TEGIPVSTILDGINIGAAIKELDDEHDPLLKKELEAKYGVLIANNLSU	1077	
1076	TAIVTSALGT	ASGFSILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETGA	1135	
1078	AAIVASIVG	AEVITFLPIAGISAGIPSLVNNELILHDKATSVVNYFNHLSSEKKYGP	1137	
1136	FTLLDPKI	IMPDDLVLSEIDFNNSITIGKCEIWRAGSGHLLTDDIDHFSSPSIY	1195	
1138	LKTEDDKIL	VPIDDLVISEIDFNNSIKLGTCNILAMEGSGHVTGNDIDHFSSPSIS	1197	
1196	RKPWLSIYD	VNLKKEKIDPSKDLMLPNAPNRFVGYEMCWTGPFSSLNDGCKLLDR	1255	
1198	HIPSLYSI	SAIGIETENLDFSKIMMLPNAPSVFWETGAVGLSLENDGTRLLDSIR	1257	
1256	DHYEGOFY	RYFAFIADALITKLPRYEDTNVINLNDGNTSRISVPIVITEQIRKNLSYS	1315	
1258	DLPGFYR	WRYAFYF-DYAITLKPYVEDINIKILDKOTRNFIMTITNEIRNKL	YSYS 1316	
1316	FYGGSGYS	LSLSPYNNMIDLNVENDTWVIDVNVVXNITISDBIOKGLIENILSKL	1375	
1317	FDGAGGYS	LLSSYPISTNINLSKODLWTFINLNEVREISIBENGITKKGKLIKDVLSKI	1376	
1376	NIBDNKI	LNNHINFGDINESRPFISLTFSELEDINIIETDLVSKYSKILLSGNCKM	1435	
1377	DINKKLI	IGNQIDFSGDIDNKDRIYFLTCELDOKLSLIEINLVAKSYLSLSDGKNY	1436	
1436	LIENSSDI	OQKIDHIGENGHEQRYPIYSYTD-NETKYNGPIDYSKXEGLFTABFSNESII	1494	
1437	LISNLSNT	IEKINTLGLD---SKNIAINYDESNNKYFGAI-----SKTSQKSII	1483	
1495	RNTYMPD	NNL-----FYSGKDL--KDIRIINKGDVKLILIGNVFKD---DMKVLSL	1541	
1484	H--YKDS	KNILEFYNDSTLEFNSKDFIAEDINVFMKDDINTITGKYVVDNNTDKSIDFS	1541	
1542	FTIEDNTI	KLVGVLDENGVAQILKEWNAKSA	LNTNSIAMFLBSINIKIFYNNLDP 1601	
1542	ISLVSKQV	KVNGHYLINESVSSYLD	FPVKNSDGHNTSNFMFLDNISFWKLFGE--- 1598	
1602	NIEPILDT	NFTIISGNSIGOFELICDKDKNIQVPIFNFKI	KETSYTLVGNRONLIVEPS 1661	
1599	NINEVIDK	YTLVGTKNLGVVEFLCDNNKNI	DIYFGEMKTSSEKSIIFSNGENVVVEPI 1658	
1662	YHLLD	SGNISTVINFSQKLYGIDRVYNNKV	IAPNLYT 1700	
1659	YN-PDTE	DISTSDSFVEPYLYGIDRVYNKVL	IAPDLYT 1696	

RESULT 5

RESULT 5
US-10-354-774-6

US-10-334-774-8
: Sequence 6. Application US/10354774

; sequence 6, APPLICATION 03/1033
; Publication No. US20030215468A1

; FURTHER INFORMATION NO: 002003
; GENERAL INFORMATION:

APPLICANT: Will

Thal

TITLE OF INVENTION: Multivalent Vaccine For Clostridium

Botulinum Neurotoxin

NUMBER OF SEQUENCES: 82

CORRESPONDENCE ADDRESS:

ADDRESSEE: Medlen & Carroll

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/354,774
FILING DATE: 30-Jan-2003
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/704,159
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPFD-02304
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
PS-10-354-774-6

Query Match 51.5%; Score 4467; DB 15; Length 2710;

Db 540 GSLSENGVDNFKNATLADKNYLLANKIPSNVVEEAGSKNYHYIQLQDDISYEATCNL 599
Qy 598 FSKOPVSSILYQKNIEGSTAYYYVAD--AEIKIDKVRIPYQISNKENIKLTFIGHGK 655
Db 600 FSKNPKNSIIQORN--NESAKSYFLSDGSEILNKRIPERLKNKEKVKVTFIGHGK 657
Qy 656 SEFNTPTANLDVDSLSSEIETILNKAADIPKPIEINLGCNMFYSISAEETYPGKL 715
Db 658 DEFNTSEPARLSVDSLSNISISFLDTIKLIDSPKVEVNLGCNMFSDPVEETYPGKL 717
Qy 716 LKIKDRVSELMPSISQDSITYSANQYEVINEEGKREILDSHGKWKINKESIIKDISK 775
Db 718 LLSIMDKITSLPDVKNKSITIGANQYEVINEEGKREILDSHGKWKINKESIIKDISK 777
Qy 776 EYISFNPKNKIIVKSKYHEILSTLQIBRNANSSDIDLEKKMLTECEINVASNDRQ 835
Db 778 EYIFDSIDNKLKAKSNIPGLASISEDIKTLDDASVSPDTKFTLNKLKNISSIGDY 837
Qy 836 IVEGREEBEAKNLTSDINIKNEFKLIESISLVDLKHQGLDSDSHFISPDISKTEG 895
Db 838 IYKELEPVKNIIHNSIDIDLEFNLENVSDELYELKKNLDEKYLISFEDISKNNST 897
Qy 896 FRIRFINKETGSIETETKEIFSEYATHISKEISNIDTIIDNVNGKLVKKNVDAAE 955
Db 898 YSVRFINKNGESVYVETEKEIFSKYSEHTKEISTIKNSIITDVNGLDNIQLDHTSQ 957
Qy 956 VNTLNSAFTIQSILYNTTKESLNSVAKVQVLAQFSTGLNTITDASKVVELVSTAL 1015
Db 958 VNTLNSAFTIQSILYNTTKESLNSVAKVQVLAQFSTGLNTITDASKVVELVSTAL 1017
Qy 1016 DETIDLLPLSEGLPIIATIDVSLGAAIKELSETNDPLRQETAKIGIVAVNLTAAS 1075
Db 1018 NDTINVLPITTEGIPVITLDGINLGAIKELDEHPLLKKEAKGVLAINNLSI 1077
Qy 1076 TAVTSALGTASGFSILLVPLAGISAGIPSLVNNELILODKATKVIDYFKHISLAETGA 1135
Db 1078 AAVTASIGVIGAEVITFLPIAGISAGIPSLVNNELIHDKATSVVYFNHLSBSKYG 1137
Qy 1136 FTLLDDKIMPODDLVLSEIDFNNSITILGKEIMRAEGSGHTLTDDIDHFSSPSITY 1195
Db 1138 LKTEDDKILVPIDDLVISEIDFNNSISIKLGTGNLAWEGSGHTVTGNIDHFSSPSISS 1197
Qy 1196 RKPWLSIYDLNITKKBKIPSKDLAVLPNAPRVFQYEMGTGPFPSLNDGTGKLDLR 1255
Db 1198 HIPSLSIYSAIGIETENLDFSKIMLPNAPS RVFWETGAVPGLRSLNDGTGRLDLSR 1257
Qy 1256 DHYEQQYRYPAFIADALITKLPYEDTNVRLNDGNTSRFIVPITTEIRKNLSYS 1315
Db 1258 DLYPGKPYRYPFAFF-DYAITTLKPYEPTNIXIKLKDQTRNFMPTITTNIRKNLSYS 1316
Qy 1316 FYGSGGSYLSLSPYNNIDNLVENDTVIDVNVKNITTESDEIQGELIENITSLK 1375
Db 1317 PDGAGGYSLLSYPISTINUSKODLWIFNIDNEVREISLNGITKKGKLIKOVLSKI 1376
Qy 1376 NIEDNKILNNHINFYGDINESNRFTISLTFISLEIDILVSKYSKILLGNCMK 1435
Db 1377 DINKKLIIGNQIDRSGDIDNDRYIFLTCELDDKISLIEINLVAKSYSLLSGDKY 1436
Qy 1436 LIENSSDIOQIDHIGBHQKYPYSYID-NETKNGFIDYSKKEGLFTAEFNSIESII 1494
Db 1437 LISNLSNTIEKINTGLD---SKNIANYTDESNNKYFGAI-----SKTSQKSI 1483
Qy 1495 RNIMYPSNNL-----FYISKOL--KDIRINKGDVKKLIGNYFKD---DMKVSLS 1541
Db 1484 H--YKDSKNILFYNDSTLEFNSKOFIABDINVMFKODINTITGKYVNDNNTDKSIDFS 1541
Qy 1542 FTIEDTNTIKLVYLDENGVAQILKFMNNAKALNTSNLNMFLSINIKIFNNLDP 1601
Db 1542 ISLVSKNOVKVNGLYLINESYSSVLDVFNKSDGHNTSNFMFLDNISFWKLFGFE--- 1598
Qy 1602 NIEFILTNTIIGSGNISIGOFELICDKORNIQPIFNFKIETSYTLYVGNRQNLIVEPS 1661

Db 1599 NINPVIDKYFTLVGKTNLGVFEICDNNKNIDIFYGKWTSSSKSTIFSGNGRNVVPEPI 1658
Qy 1662 YHLDDSNISSTISQKYLIGIDRYVKNKVIIPNLYT 1700
Db 1659 YN-PDGTGEDISTSLDSYEPGLYIGIDRYINKVLIAPDLYT 1696
RESULT 6
US-10-271-012-6
; Sequence 6, Application US/10271012
; Publication No. US20030219457A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/271,012
; FILING DATE: 15-Oct-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-271-012-6
Query Match 51.5%; Score 4467; DB 15; Length 2710;
Best Local Similarity 50.7%; Pred. No. 9.8e-271;
Matches 871; Conservative 366; Mismatches 440; Indels 42; Gaps 15;
Qy 1 MNLVKKQLQKQVYVFKRIQDEYVAIINALVEEYHNMSSESVVEKYLKLDKNNLTNYL 60
Db 1 MSLSKELIKLAY-SIRPRENEVYKTLITLNDLYNNENKYLQKKNESIDVFM 59
Qy 61 NTKYKSGNKKALKKFKKYLTVLEVKLNNSLTPVEKNLHFVIGGQINDTAINYNQWD 120
Db 60 NKYKTSSENRLSKKDLKEVLIIKNSNTPSPVEKNLHFVWIGGVSIDALEYIKQWAD 119
Qy 121 VNSDYTKVYVDSNAFLINTLKKTIIVESATNTLTSEFRENLDPEFDYKNFKYKRMETIY 180
Db 120 INAEYTKLWYDSFAFLVNTLKKAVSSSTTEALQLLEEEIQFONMKFYKRMETIY 179
Qy 181 DKQHFIDYKQSQEENPEFTIDNIITKLYSNEYSKDLALNKYIEESLNKITANNNGDI 240
Db 180 DRQKRFINYKSLQNKPTVPTIDDIKSHLVSEINRDETLYESVTSLSRKNINHGDI 239
Qy 241 RNLEKFADEDLVRLYNQBELVERWNLAASDILRISMLKEDGGVYLDVILPGIQDLPKS 300

Db 240 RANSLTEQELNINYOELLNRGNAASDIIVALLAKNFGVYLDVMDLPGHSDLPKT 299
Qy 301 INKPDSTINTSWEMIKLEAMKYKEVPGVTSKNFMDLBEVQORSFESALSXSDEIF 360
Db 300 ISRPSSIGLDRWEMIKLEAMKYKYNNTYSENFPKLDQKDNFKLIIESKSEKSEIF 359
Qy 361 LPDDDKVSPLEVKIAPANNVINOALISLKDSCDIAVINOIKRYKILNDMLNPSINE 420
Db 360 SKLENVSDLEIXIAFALGVSINOALISKQGSYLTNLVIEQVKNYQFLNOHLAPAES 419
Qy 421 GTDFNTMKFSDKLAGISNEDNMFMKILNTLVKGFAPDVRSTINLSGPGVYTGAYOD 480
Db 420 DNNFTDTTKLPHSLFNASAEHSMFLTKAPLVQGFMPPEARSTLSLSPGAYASAYD 479
Qy 481 LLAFKDNSTNHILLEPRLNFPKTKISQLTQETITSLWFSNOARAKSQFBEYKGYPE 540
Db 480 FINLOENTIEKTLKASDLIEFKPPENLSQLTQETINSLWSPQAGAKQFEKYVRYDTG 539
Qy 541 GALGEDNLOFAQNTVLDKDY-VSKILSS--MKTRNKEYIHVIVOLQGDKISYEASCNL 597
Db 540 GSLSEDNGVDFNKNTALDKNYLLNNKIPSNVVEAGSKNYVHIIQLOGDDISYEATCNL 599
Qy 598 FSKDPYSILYQKIEGSETAYVYVAD--AEIKEIDKYRIPYOISNKRNIKLTFIGHGK 655
Db 600 FSKNPKNSIIQRMN--NESAKSYFLSDGSESILELNKRIPELKNKEKVKVTFIGHGK 657
Qy 656 SEFNTDFANLVDLSSETETIILAKADISPKYIEINLLGCNMFYSISABETYPGKL 715
Db 658 DEFNTSEFARLSVDLSNEISSFLDTKLDISPKVEVNLGCNMFSDYDNVEETPGKL 717
Qy 716 LKIKDRVSELMPSIODSITVSANQYEVNRINEGKEIILDHSGKNINKEESIIKDISK 775
Db 718 LLSIMDKITSLPDWNKNSITIGANQYEVNRINEGKELLARHSGKNINKEEAIMSLSSK 777
Qy 776 BYISFNPKENKIIVKSVLHELSTLLQEIERNANSSDIDLEKKVMTCECINVASNDRQ 835
Db 778 EYIFPDSIDNKLKAKSKNIPGLASISEDITLLLDASVSPDTKPIILNNLKNLTISSIGDY 837
Qy 836 IVEGRIBEAKNLSDSINVYKNEFKLIESDSYLDLKHQGLDSDSHFISFEDISKTENG 895
Db 838 IYBEKLEPVKNIIHNSIDDLDEFNLLENVDELYELKLNLDKELIISFEDISKNST 897
Qy 896 FRIRPINKETGNSPIETKEIFSEYATHISKEISNIKOTIFDNVNGKLVKNVNLDAHE 955
Db 898 YSVRFINKNGESVYVETEKEIFSKYSEHITKEISTIKNSIITDVNGLLDNIQLDHTSQ 957
Qy 956 VNTLNSAFFIQLSEYNTKESLNLVAMKVQYVLAQFSLTGLNTIITDASKVVELVSTAL 1015
Db 958 VNTLNAAPFIQSLIDYSNKNVDLSTSVKQVLAQFSLTGLNTIYDSIQVNLNLSAV 1017
Qy 1016 DETIDLLPLTSLGLPIIATIIDGVSIGAAIKELSETNDPILROEIAKIGIAMVNLTAAS 1075
Db 1018 NDTINVLPTITEGIVSTIIDLGINLGAALKELDEHDPDLKKELEAKGVGLIANNLSI 1077
Qy 1076 TAIVTSALCIAGSGFILLVPLAGISAGIPSLVNNELIILQDKATKVIDYFKHISLAETEGA 1135
Db 1078 RAIVASIVIGIAGEVTFILPFIAGISAGIPSLVNNELIILHDKATSVNRYNHLSSKKYGP 1137
Qy 1136 FTLLDDKIIMPODDVLVSEIDFNNSITLKGCEIWRAGSGGSHLTDDIDHFFSPSPITY 1195
Db 1138 LXTEDDKILVPIDDLVISEIDFNNSIKLGTGNILAMEGGSGHTVTGNIDHFFSPSPISS 1197
Qy 1196 RKPWLSIYVNLNIKEKIDFSDKLMVLPNAPNRVFGYEMGTWPGPSRLDNDGTLLDRIR 1255
Db 1198 HIPSLSIYGAIGIETENLDFSKIMMLPNAPSRVFWETGAVPGFLRSLENDGTLLDRIR 1257
Qy 1256 DVEGOFYWRYPAFIADALITIKLKPREDTNRVIRNLGDNTRSFIVPVIITTEQIRKNLSYS 1315
Db 1258 DLTGPKFYWRFYAFF-DYAITLTKPVIEDTNIKIKDKOTRNFNFTITTEIRKNLSYS 1316
Qy 1316 FYGGSGYSLSISPPYNNIDNLVENDTWTVIDVNVKNITTESDEIQGELIENILSKL 1375

Db 1317 FDGAGGYTSLLSYPSTNINLSKODLMIFNIDNEVRISIENTGIIKKGLIKDLVLSKI 1376
Qy 1376 NIEDNKIILNNHINFYGDINESNRFTISLTFSILEDINIIIEIDLVSXKYLKSGNCKM 1435
Db 1377 DINKKLIIGNOTIDFSGDIDKNDRIFLTCELDKISLIIIEINLVAKSYSLLSGDKNY 1436
Qy 1436 LIENSIDIQKIDHIGFNGHQKIPYSYID-NETKNGFIDYSKKEGLEFASFSNESII 1494
Db 1437 LISLSNTIEKINTGLD--SKNIAINYTDESNNKYFGAI-----SKTSQKSI 1483
Qy 1495 RNIMPSNNL-----FIYSKDL--KDIRINKGDVKKLLIGNYFKD---DMKVSLS 1541
Db 1484 H--YKDSKNILEFYNDSTLEFNSKOFIAEDINVMKODINTITGKYVDNNTDKSDFDS 1541
Qy 1542 FTIEDTNTIKLVYLDENGVAQILKFMNNAKALANTSLSLMMFLESINIKNIFNNLDP 1601
Db 1542 ISLVSKQKVKNGLYNESVSSVLDLVKNSDGHNTSNFNNFLDNLISFWKLFGE-- 1598
Qy 1602 NIEFILTNTIISGNSIGOFELICDKDKNIQPIFNPKIKETSITLYVGNRQNLIVEPS 1661
Db 1599 NINFVIDKYFTLVGKTNLGVYEFICDNNKIDIVFGWKTSKSSKSTIFSGNGRNVVVEPI 1658
Qy 1662 YHLDOSGNISSTVNFQKLYGIDRYNVKVIAPNLAT 1700
Db 1659 YN-PDTGSDISTSLDFSEPLYGIDRYNKVLIAPDDYT 1696

RESULT 7
US-10-463-957-1
; Sequence 1, Application US/10463957
; Publication No. US20040028705A1
; GENERAL INFORMATION:
; APPLICANT: Ballard, Jimmy D.
; APPLICANT: Spyles, Lea M.
; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637
; CURRENT APPLICATION NUMBER: US/10/463,957
; CURRENT FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: 60/389,685
; PRIOR FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 556
; TYPE: PRN
; ORGANISM: Clostridium difficile
US-10-463-957-1

Query Match 25.4%; Score 2204; DB 12; Length 556;
Best Local Similarity 75.0%; Pred. No. 8.5e-130;
Matches 417; Conservative 66; Mismatches 73; Indels 0; Gaps 0;

Qy 1 MNLVNAQIQKQVYKVFRIQDEYVAILNALFEYHNMSESVVEKYKLKDINNLTNDYL 60
Db 1 MSLVNRKQLEKMANVRFRTQDEYVAILDALFEYHNMSESVVEKYKLKDINNLTNDYL 60
Qy 61 NTYKSGRNKALKKPKXYLTMTVELEKNSLTPVEKNLHFTWIGGINDTAINYNQWKD 120
Db 61 DTYKSGRNKALKKPKXYLTMTVELEKNSLTPVEKNLHFTWIGGINDTAINYNQWKD 120
Qy 121 VNSDYTVKYFVPSNAFLINTLTKTIVESATNNTLSFRENLDNPFEDYNKFKRMEIY 180
Db 121 VNSDYNNVYFVPSNAFLINTLTKTIVESAINNTLSFRENLDNPFEDYNKFKRMEIY 180
Qy 181 DKQKHFDVYKSOIENPEFIIDNIKTLYSNEYSKDLEALNKYIEESLNKITAANGNDI 240
Db 181 DKQKNINYYKAGREENPELLIDDIKTYLSNEYSKEIDELATYIEESLNKITAANGNDV 240
Qy 241 RNLEKFADEDLVLYNQELVERWNLAASDILRISMLKEDGGVYLDVLDILPGIOPDLFKS 300
Db 241 RNFEFEKNGSFLYEQELVERWNLAASDILRISALKETGGMYLDVMDLPGIOPDLFES 300
Qy 301 INKPDSTINTSWEMIKLEAMKYKEVPGVTSKNFMDLBEVQORSFESALSXSDEIF 360


```
Db 301 IEKPSVTVDFWEMTKLEAIMKYKEYIPEYTSSEHFDMLDEEVQSFESVLASKDKSEIF 360
Qy 361 LPDDIKVSPLEVKIAFANNVINOALISLKDSCDVLINQIKVRYKILNDNLNPSINE 420
Db 361 SSLGDWEASPLEVKIAFNKGIINOGLISVKDSCNLIQVQIENRYKILNNSLPAISE 420
Qy 421 GTDFNTMTKIFDCKLASINEDNMFMKITYLKYGFAPDVRSTINLSGPGVYTGAYOD 480
Db 421 DNDFTNTTTFIDSIMAEANADNGRFMMELGKYLVRGFPDPVKTITNLSGPEAYAAAYOD 480
Qy 481 LLMFKDNSTNHLLEPELRNFPFKTKISQLEQITSLWSFNQARAKSQFEYKKGYPF 540
Db 481 LLMFKEGSNMHLIEADLRNFEISKTNISQTEQEMASLWSFDDARAKAQFEYKKNYFE 540
Qy 541 GALGEDDNLDPQNTV 556
Db 541 GSLGEDDNLDFSQNV 556

RESULT 8
US-10-463-957-11
; Sequence 11, Application US/10463957
; Publication No. US20040028705A1
; GENERAL INFORMATION:
; APPLICANT: Ballard, Jimmy D.
; APPLICANT: Spyres, Lea M.
; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637
; CURRENT APPLICATION NUMBER: US/10/463,957
; CURRENT FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: 60/389,685
; PRIOR FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 11
; LENGTH: 556
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Completely synthesized
; NAME/KEY: MISC FEATURE
; LOCATION: (395)..(395)
; OTHER INFORMATION: Xaa at position 395 is ala, asp, glu, phe, gly, his, ile, lys,
; OTHER INFORMATION: leu, met, asn, pro, gln, arg, ser, thr, val, trp, or tyr.
US-10-463-957-11

Query Match 25.3%; Score 2193; DB 12; Length 556;
Best Local Similarity 74.8%; Pred. No. 4.2e-129;
Matches 416; Conservative 66; Mismatches 74; Indels 0; Gaps 0;

Qy 1 MNLVNAQLQKQVYVKFRIQDEYVAIINALEEYHNMSSESVVEKYLKDKDINNLTNYL 60
Db 1 MSLVNRKQLEKMANVRPRTQDEYVAIILDALEEYHNMSNTVVEKYLKDKDINSLTDIYI 60
Qy 61 NTKYKSGRNKALKKPKFVLTMEVLELKNKSLTPVEKNLHPFIWIGQINDTAINYNQWKD 120
Db 61 DTYKSGRNKALKKPKFVLTVEVLELKNKSLTPVEKNLHPFIWIGQINDTAINYNQWKD 120
Qy 121 VNSDVTVKVYDSNAFLNTLTKTIVESATNTLESFRENLDNPFDFYKFKRMEIY 180
Db 121 VNSDYNVNVFYDSNAFLNTLTKTIVESAINDTLESFRENLDNPFDFYKFKRMEIY 180
Qy 181 DKQKHFIDYKSOIENPEFIIDNIITKTYLSNEYSKOLEALNKYIEESLNKITTANNNDI 240
Db 181 DKQKFINFYKQARENPELIIDDIIVKTYLSNEYSKEIDELNTYIEESLNKITQSGNDV 240
Qy 241 RNLEKFADEDLVRLYNQELVERWNLAASDILRISMLKEDGGVYLDVILPGIQDPLFKS 300
Db 241 RNFEFKNGESFNLYEQELVERWNLAASDILRISALKKEIGMYLDVDMPLGQDPLFES 300
Qy 301 INKPSITNTSWEMTKLEAIMKYKEYIPGYTSKNFDMDEEVQSFESALSKSKSEIF 360
Db 301 IEKPSVTVDFWEMTKLEAIMKYKEYIPEYTSSEHFDMLDEEVQSFESVLASKDKSEIF 360
Qy 361 LPDDIKVSPLEVKIAFANNVINOALISLKDSCDVLINQIKVRYKILNDNLNPSINE 420
Db 361 SSLGDWEASPLEVKIAFNKGIINOGLISVKDSCNLIQVQIENRYKILNNSLPAISE 420
Qy 301 INKPSITNTSWEMTKLEAIMKYKEYIPGYTSKNFDMDEEVQSFESALSKSKSEIF 360
```

```
Db 301 IEKPSVTVDFWEMTKLEAIMKYKEYIPEYTSSEHFDMLDEEVQSFESVLASKDKSEIF 360
Qy 361 LPDDIKVSPLEVKIAFANNVINOALISLKDSCDVLINQIKVRYKILNDNLNPSINE 420
Db 361 SSLGDWEASPLEVKIAFNKGIINOGLISVKDSCNLIQVQIENRYKILNNSLPAISE 420
Qy 421 GTDFNTMTKIFDCKLASINEDNMFMKITYLKYGFAPDVRSTINLSGPGVYTGAYOD 480
Db 421 DNDFTNTTTFIDSIMAEANADNGRFMMELGKYLVRGFPDPVKTITNLSGPEAYAAAYOD 480
Qy 481 LLMFKDNSTNHLLEPELRNFPFKTKISQLEQITSLWSFNQARAKSQFEYKKGYPF 540
Db 481 LLMFKEGSNMHLIEADLRNFEISKTNISQTEQEMASLWSFDDARAKAQFEYKKNYFE 540
Qy 541 GALGEDDNLDPQNTV 556
Db 541 GSLGEDDNLDFSQNV 556

RESULT 9
US-10-463-957-12
; Sequence 12, Application US/10463957
; Publication No. US20040028705A1
; GENERAL INFORMATION:
; APPLICANT: Ballard, Jimmy D.
; APPLICANT: Spyres, Lea M.
; TITLE OF INVENTION: MUTANTS OF CLOSTRIDIUM DIFFICILE TOXIN B AND METHODS OF USE
; FILE REFERENCE: 5820.637
; CURRENT APPLICATION NUMBER: US/10/463,957
; CURRENT FILING DATE: 2003-06-17
; PRIOR APPLICATION NUMBER: 60/389,685
; PRIOR FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 12
; LENGTH: 556
; TYPE: PRT
; ORGANISM: Clostridium difficile
; OTHER INFORMATION: Clostridium difficile
US-10-463-957-12

Query Match 25.2%; Score 2190; DB 12; Length 556;
Best Local Similarity 74.8%; Pred. No. 6.5e-129;
Matches 416; Conservative 66; Mismatches 74; Indels 0; Gaps 0;

Qy 1 MNLVNAQLQKQVYVKFRIQDEYVAIINALEEYHNMSSESVVEKYLKDKDINNLTNYL 60
Db 1 MSLVNRKQLEKMANVRPRTQDEYVAIILDALEEYHNMSNTVVEKYLKDKDINSLTDIYI 60
Qy 61 NTKYKSGRNKALKKPKFVLTMEVLELKNKSLTPVEKNLHPFIWIGQINDTAINYNQWKD 120
Db 61 DTYKSGRNKALKKPKFVLTVEVLELKNKSLTPVEKNLHPFIWIGQINDTAINYNQWKD 120
Qy 121 VNSDVTVKVYDSNAFLNTLTKTIVESATNTLESFRENLDNPFDFYKFKRMEIY 180
Db 121 VNSDYNVNVFYDSNAFLNTLTKTIVESAINDTLESFRENLDNPFDFYKFKRMEIY 180
Qy 181 DKQKHFIDYKSOIENPEFIIDNIITKTYLSNEYSKOLEALNKYIEESLNKITTANNNDI 240
Db 181 DKQKFINFYKQARENPELIIDDIIVKTYLSNEYSKEIDELNTYIEESLNKITQSGNDV 240
Qy 241 RNLEKFADEDLVRLYNQELVERWNLAASDILRISMLKEDGGVYLDVILPGIQDPLFKS 300
Db 241 RNFEFKNGESFNLYEQELVERWNLAASDILRISALKKEIGMYLDVDMPLGQDPLFES 300
Qy 301 INKPSITNTSWEMTKLEAIMKYKEYIPGYTSKNFDMDEEVQSFESALSKSKSEIF 360
Db 301 IEKPSVTVDFWEMTKLEAIMKYKEYIPEYTSSEHFDMLDEEVQSFESVLASKDKSEIF 360
Qy 361 LPDDIKVSPLEVKIAFANNVINOALISLKDSCDVLINQIKVRYKILNDNLNPSINE 420
Db 361 SSLGDWEASPLEVKIAFNKGIINOGLISVKDSCNLIQVQIENRYKILNNSLPAISE 420
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 16:37:55 ; Search time 63.75 Seconds

(without alignments)
8413.819 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1700

Perfect score: 8675

Sequence: 1 MNLVKAQLQKVVYKFRIQ.....YLYGIDRYKVKVYIAPNLYT 1700

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPTRMBL 25:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_rhnc:*
9: sp_organelle:*
10: sp_phase:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*
16: sp_rvirus:*
17: sp_bacteriap:*
18: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8675	100.0	2364	2 Q46342	Q46342 clostridium
2	6787	78.2	2367	2 Q9EXR0	Q9EXR0 clostridium
3	6758	77.9	2367	2 Q9F931	Q9F931 clostridium
4	6737	77.7	2367	2 Q46034	Q46034 clostridium
5	2156.5	24.9	554	2 Q93L39	Q93L39 clostridium
6	2082.5	24.0	2178	2 Q46149	Q46149 clostridium
7	1916	22.1	698	2 Q68553	Q68553 clostridium
8	1898	21.9	698	2 Q9EXQ8	Q9EXQ8 clostridium
9	1892	21.8	697	2 Q86141	Q86141 clostridium
10	1474	17.0	553	2 Q9FCX5	Q9FCX5 clostridium
11	512.5	5.9	3169	2 Q9ZGR4	Q9ZGR4 escherichia
12	507.5	5.9	3169	2 Q82316	Q82316 escherichia
13	494	5.7	3317	16 Q8SWP8	Q8SWP8 mycoplasma
14	489	5.6	2792	5 Q814R2	Q814R2 plasmodium
15	479	5.5	2771	5 Q26216	Q26216 plasmodium
16	468.5	5.4	2965	5 Q8WRS5	Q8WRS5 plasmodium

17 467 5.4 2752 5 Q9BJY0 Q9BJY0 plasmodium
18 461 5.3 2976 5 Q8WRS4 Q8WRS4 plasmodium
19 460 5.3 5767 5 Q81525 Q81525 plasmodium
20 454.5 5.2 2957 5 Q8WRS6 Q8WRS6 plasmodium
21 448.5 5.2 3081 5 Q7YWF1 Q7YWF1 plasmodium
22 448.5 5.2 3203 5 Q7YWF7 Q7YWF7 plasmodium
23 448 5.2 3130 5 Q9BK46 Q9BK46 plasmodium
24 448 5.2 3130 5 Q81DX6 Q81DX6 plasmodium
25 448 5.2 3254 5 Q9BK45 Q9BK45 plasmodium
26 448 5.2 3254 5 Q7YWF9 Q7YWF9 plasmodium
27 447.5 5.2 3080 5 Q7YWF0 Q7YWF0 plasmodium
28 447.5 5.2 3203 5 Q7YWF6 Q7YWF6 plasmodium
29 447 5.2 3096 5 Q7YWF2 Q7YWF2 plasmodium
30 447 5.2 3256 5 Q7YWF8 Q7YWF8 plasmodium
31 440 5.1 2269 5 Q26223 Q26223 plasmodium
32 439.5 5.1 2770 5 Q7YUE9 Q7YUE9 plasmodium
33 436.5 5.0 2184 5 Q81UQ6 Q81UQ6 plasmodium
34 435.5 5.0 2747 5 Q9BJX9 Q9BJX9 plasmodium
35 433.5 5.0 2849 5 Q81HV4 Q81HV4 plasmodium
36 431.5 5.0 4688 16 Q9PQ08 Q9PQ08 ureaplasma
37 422 4.9 2166 16 Q51465 Q51465 borrelia bu
38 418 4.8 3381 5 Q812V4 Q812V4 plasmodium
39 412.5 4.8 1956 5 Q81IE1 Q81IE1 plasmodium
40 406.5 4.7 6761 5 Q81C77 Q81C77 plasmodium
41 402 4.6 1686 5 Q81RP9 Q81RP9 plasmodium
42 401.5 4.6 3223 2 Q8L1L9 Q8L1L9 escherichia
43 399.5 4.6 1711 5 Q8MWP2 Q8MWP2 plasmodium
44 399.5 4.6 3223 2 Q9RPH1 Q9RPH1 escherichia
45 399 4.6 3724 5 Q77320 Q77320 plasmodium

ALIGNMENTS

RESULT 1
Q46342
ID Q46342 PRELIMINARY; PRT; 2364 AA.
AC Q46342;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Cytotoxin L.
OS Clostridium sordellii.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1505;
RN [1]_
RC SEQUENCE FROM N.A.
RC STRAIN=6018;
RX MEDLINE=95369733; PubMed=7642137;
RA Green G.A., Schue V., Montell H.;
RT "Cloning and characterization of the cytotoxin L-encoding gene of
RT Clostridium sordellii: homology with Clostridium difficile cytotoxin
RT B.";
RL Gene 161:57-61(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=6018;
RX MEDLINE=96149194; PubMed=8544213;
RA Green G.A., Schue V., Girardot R., Montell H.;
RT "Characterisation of an enterotoxin-negative, cytotoxin-positive
RT strain of Clostridium sordellii.";
RL J. Med. Microbiol. 44:60-64(1996).
DR EMBL; X82638, CAA57959.1; -;
DR FIR; I40884; I40884.
DR GO; GO:0003743; F:translation initiation factor activity; IEA.
DR GO; GO:0006413; P:translational initiation; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR007577; Gly_transf_sug.
DR InterPro; IPR001950; TUF_SUI1.
DR Pfam; PF01473; CW_binding_1; 18.
DR Pfam; PF04486; Gly_transf_sug; 1.
DR PROSITE; PS01118; SUI1_1; 1.

SQ	SEQUENCE	2364 AA; 270576 MW; EAD8A4467A89BDBB CRC64;
	Query Match	100.0%; Score 8675; DB 2; Length 2364;
	Best Local Similarity	100.0%; Pred. No. 0;
	Matches 1700; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1	MNLVYKQLOQWVVKFRIQSDVVAIIINALEEYHNMSESSVEKYLKLDIKNLTDNYL 60
DB	1	MNLVYKQLOQWVVKFRIQSDVVAIIINALEEYHNMSESSVEKYLKLDIKNLTDNYL 60
QY	61	NTYKSGRNKALKKFKFKEVLTWEVLELKNLSLTPVEKNLHFIWIGGQINDTAIYNQWKD 120
DB	61	NTYKSGRNKALKKFKFKEVLTWEVLELKNLSLTPVEKNLHFIWIGGQINDTAIYNQWKD 120
QY	121	VNSDYTVKVFVDSNAFLINTLKTIIVESAATNTLTSPRENLDPEFDYNYKFKRMEIYY 180
DB	121	VNSDYTVKVFVDSNAFLINTLKTIIVESAATNTLTSPRENLDPEFDYNYKFKRMEIYY 180
QY	181	DKQHFIDYKQSQJEEENPEFIIDNIITKYLSEYKSDLEALNKYIEESLNKIITANNNDI 240
DB	181	DKQHFIDYKQSQJEEENPEFIIDNIITKYLSEYKSDLEALNKYIEESLNKIITANNNDI 240
QY	241	RNLEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVDILPGIQPOLFKS 300
DB	241	RNLEKFADEDLVRLYNQELVERNLAAASDILRISMLKEDGGVYLDVDILPGIQPOLFKS 300
QY	301	INKPDSINTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEVOQSPFSALESKSDKSEIF 360
DB	301	INKPDSINTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEVOQSPFSALESKSDKSEIF 360
QY	361	LPLDDIKVSPLEVKIAPANNVINQALISLSDSCSCLVINQIKRYKIINDNLNPSINE 420
DB	361	LPLDDIKVSPLEVKIAPANNVINQALISLSDSCSCLVINQIKRYKIINDNLNPSINE 420
QY	421	GTDFNTTWKIFSDKLASISNEDNMFMKIITNYLKVGFADPVRSTINLSGPGVYTGAYQD 480
DB	421	GTDFNTTWKIFSDKLASISNEDNMFMKIITNYLKVGFADPVRSTINLSGPGVYTGAYQD 480
QY	481	LLMFKDNSTNHLLEPELNEFPFKTKISQTEQIITSLMSPNOARAKSQFEEYKKGYPE 540
DB	481	LLMFKDNSTNHLLEPELNEFPFKTKISQTEQIITSLMSPNOARAKSQFEEYKKGYPE 540
QY	541	GALGEDDNLDPQNTVLDKQVSKKILSSMTRNKEYIHYIQLQGDKIISYEASCNLFSK 600
DB	541	GALGEDDNLDPQNTVLDKQVSKKILSSMTRNKEYIHYIQLQGDKIISYEASCNLFSK 600
QY	601	DPYSILYQKNIEGSETAYYYVADAEIKETDKYRIIPYOISNKENIKLFTIGHGKSEFNT 660
DB	601	DPYSILYQKNIEGSETAYYYVADAEIKETDKYRIIPYOISNKENIKLFTIGHGKSEFNT 660
QY	661	DTFANLDVDSLSETETTLNLAKADISPKYTEINLLGCNMPYSISABETYPGKLILKIK 720
DB	661	DTFANLDVDSLSETETTLNLAKADISPKYTEINLLGCNMPYSISABETYPGKLILKIK 720
QY	721	DRVSELMPSISQDSITVSANQYEVRIINEEGKREILDHSGKWINKESIIKDISKEYISF 780
DB	721	DRVSELMPSISQDSITVSANQYEVRIINEEGKREILDHSGKWINKESIIKDISKEYISF 780
QY	781	NPKENKIIIVKSKYLHELSTLLOETRNNANSSDIDLEKKVMLTECEINVASNIDRQIVEGR 840
DB	781	NPKENKIIIVKSKYLHELSTLLOETRNNANSSDIDLEKKVMLTECEINVASNIDRQIVEGR 840
QY	841	IEEAKNLTSDSINYIKNFKLIESISDSLYDLKHONGLDSDHFTSFEDISKTEGPRIRF 900
DB	841	IEEAKNLTSDSINYIKNFKLIESISDSLYDLKHONGLDSDHFTSFEDISKTEGPRIRF 900
QY	901	INKETGNSIFITEKEIFESEVATHISKEISNIKOTIFDNVNGKLVKKVNLDAAEVNTLN 960
DB	901	INKETGNSIFITEKEIFESEVATHISKEISNIKOTIFDNVNGKLVKKVNLDAAEVNTLN 960
QY	961	SAFFTQSIEYNTTKESISNLVAMKVQVYIQLFSTGLNTITDASKVVELVSTALDETID 1020
DB	961	SAFFTQSIEYNTTKESISNLVAMKVQVYIQLFSTGLNTITDASKVVELVSTALDETID 1020

QY	1021	LLPTLSEGLPIATIIDGVSIGAAIKELSETNDPLLQREIAKIGIMAVNLTAATAIYT 1080
DB	1021	LLPTLSEGLPIATIIDGVSIGAAIKELSETNDPLLQREIAKIGIMAVNLTAATAIYT 1080
QY	1081	SALGIASGFSILLVPLAGISAGIPSLVNNELILOQKATKVIDYFKHISLAETEGAFILLD 1140
DB	1081	SALGIASGFSILLVPLAGISAGIPSLVNNELILOQKATKVIDYFKHISLAETEGAFILLD 1140
QY	1141	DKIIMPQDVLSEIDFNNNSITLKGCEIWPABEGSGHTLTDIDHDFSSPSITYRKPW 1200
DB	1141	DKIIMPQDVLSEIDFNNNSITLKGCEIWPABEGSGHTLTDIDHDFSSPSITYRKPW 1200
QY	1201	STYDVNTIKKEKIDFSKDLVLPNAPNRVFGYEWGTPGFRSLNDGDKLDRDRHYEG 1260
DB	1201	STYDVNTIKKEKIDFSKDLVLPNAPNRVFGYEWGTPGFRSLNDGDKLDRDRHYEG 1260
QY	1261	QFYWRYFAFIADALITKLKPRYEDTNVRLNDGTRSFIVPVIITTEQIRKNLSYFYGSG 1320
DB	1261	QFYWRYFAFIADALITKLKPRYEDTNVRLNDGTRSFIVPVIITTEQIRKNLSYFYGSG 1320
QY	1321	GSYLSLSFPYNNIDLVENDTWVIDVNVVKVITTESDEIQKGELIENTILSKLINIEDN 1380
DB	1321	GSYLSLSFPYNNIDLVENDTWVIDVNVVKVITTESDEIQKGELIENTILSKLINIEDN 1380
QY	1381	KIILNNHTINFVGDINESNRISLTFSLIEDINIIIEIDLVSYSKILLSGNCMKLIENS 1440
DB	1381	KIILNNHTINFVGDINESNRISLTFSLIEDINIIIEIDLVSYSKILLSGNCMKLIENS 1440
QY	1441	SDIOQKIDHIGFNGSHQKVIPIYSYIDNETKYNGFIDYSKKEGLTAEPFSNESIIRNIYMP 1500
DB	1441	SDIOQKIDHIGFNGSHQKVIPIYSYIDNETKYNGFIDYSKKEGLTAEPFSNESIIRNIYMP 1500
QY	1501	DSNNLIYSSKDLDIRIINKGDKVLLIGNYFKODMKVLSLFTTIEDNTIKNGVYLDEN 1560
DB	1501	DSNNLIYSSKDLDIRIINKGDKVLLIGNYFKODMKVLSLFTTIEDNTIKNGVYLDEN 1560
QY	1561	GVAQILKFMNNAKSALNTSNSLMFLESINIKNIFANNLDPNIBFILDNTIISGSNSIG 1620
DB	1561	GVAQILKFMNNAKSALNTSNSLMFLESINIKNIFANNLDPNIBFILDNTIISGSNSIG 1620
QY	1621	QFELICDCKDKNIQPIFNPKIKETSYTLVGNRQNLIVERPSYHLLDSDGNISSVTINFSQK 1680
DB	1621	QFELICDCKDKNIQPIFNPKIKETSYTLVGNRQNLIVERPSYHLLDSDGNISSVTINFSQK 1680
QY	1681	YLYGIDRYVNVKVIAPNLTYT 1700
DB	1681	YLYGIDRYVNVKVIAPNLTYT 1700
RESULT 2		
QSEXRO PRELIMINARY; PRT; 2367 AA.		
ID	QSEXRO	01-MAR-2001 (Tremblrel. 16, Created)
AC	QSEXRO	01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT	01-MAR-2001	01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE	Toxin B.	TCDB.
GN	Clostridium difficile.	Clostridium difficile.
OS	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;	
OC	Clostridium.	
OX	NCBI_TaxID=1496;	
RN	SEQUENCE FROM N.A.	
RC	STRAIN=8864;	
RA	von Eichel-Streiber C.;	
RL	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=8864;	
RA	Kohl M.;	
RT	"The pathogenicity locus of Clostridium difficile, strain 8864."	

NCBI_TaxID=1496;
SEQUENCE FROM N.A.
MEDLINE=20448897; PubMed=10992443;
RA Sambol S.P., Merriam M.M., Lively D., Gerding D.N., Johnson S.;
RT "Toxin gene analysis of a variant strain of clostridium difficile that
causes human clinical disease."
RL Infect. Immun. 68:5480-5487(2000).
DR EMBL; AF217292; AAC18011.1; -
DR PIR; A27636; A27636.
DR GO; GO:0003743; P:translation initiation factor activity; IEA.
DR GO; GO:0006413; P:translational initiation; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR001950; TIF_SUI1.
DR Pfam; PF01473; CW_binding_1; 17.
DR Pfam; PF04488; Gly_transf_sug; 1.
DR PROSITE; PS01118; SUI1_1; 1.
SQ SEQUENCE 2367 AA; 269337 MW; D5E715E5BD41E2F CRC64;

Query Match 77.9%; Score 6758; DB 2; Length 2367;
Best Local Similarity 76.1%; Pred. No. 2.8e-248;
Matches 1295; Conservative 204; Mismatches 200; Indels 2; Gaps 2;

QY 1 MSLVNAQLOKQVYKRIQDEVAIINALEEYHNMSSEVVEKYIKLDINLTDNYL 60
DB 1 MSLVNRKQLEKXANVRFRVQDEVAIIDLDALEEYHNMSSENTVVEKYIKLDINSLDTYI 60

QY 61 NTKYKSGNKALKKFKFVLTWEVLKXNSLTPVEKNLHFWIGGOINDTAINYINQWKD 120
DB 61 DTYKSGNKALKKFKFVLTWEVLKXNSLTPVEKNLHFWIGGOINDTAINYINQWKD 120

QY 121 VNSDYTVKVFDSNAFLINTLTKTIVESATNTLTESFRENLDNDEFFYKFRMEIYY 180
DB 121 VNSDYTVKVFDSNAFLINTLTKTIVESATNTLTESFRENLDNDEFFYKFRMEIYY 180

QY 181 DKQKHFIDYKQSENEPEFIIDNIITKYLSEYKOLEALNKYIEBSLANKIPANNNDI 240
DB 181 DKQNFQINYYKQSENEPDLDIDIVKTYLSEYKOLEALNKYIEBSLANKIPANNNDI 240

QY 241 RNLEKFADEDLVRLNQBLVERWNLAAASDLIRLSMLKEDGGVLDVLDIIPGQDPLFKS 300
DB 241 RNLEKFADEDLVRLNQBLVERWNLAAASDLIRLSMLKEDGGVLDVLDIIPGQDPLFKS 300

QY 301 INKPDQI-TNTSWEMKLEATMKYKYPGYTSKNEFMDLDEEVQSFESALSKSDKSEI 359
DB 301 INKPDQI-TNTSWEMKLEATMKYKYPGYTSKNEFMDLDEEVQSFESALSKSDKSEI 359

QY 360 FLPLDDIKVPLEVIAFANNSVINOALISLKDSCSLVINOINQIRYKILNDNLNPSIN 419
DB 361 FLPLGDIKVPLEVIAFAGKSIINQALISAKDSCSLVINOINQIRYKILNDNLNPSIN 420

QY 420 EGTDFNTMTKIPSDKLASISNDDNMFMKITYNLKVGFAFDVASTINLSPGVYTGAYQ 479
DB 421 QGNDFNTMTNFGESIGAIANENISFIAKIGSYLRVGFPEANTVTVLSGFTIYAGAYK 480

QY 480 DILLMKDSTNHLLEPELRNPEPKTKISQTEQITSLASFQARAKSOFEEYKGYF 539
DB 481 DILLMKDSTNHLLEPELRNPEPKTKISQTEQITSLASFQARAKSOFEEYKGYF 540

QY 540 EGALGEDDNLDAQNTVLDKDVVSKILSSMKTRNKEYIHYIVQLQGIKISYEASCNLFS 599
DB 541 EGALGEDDNLDAQNTVLDKDVVSKILSSMKTRNKEYIHYIVQLQGIKISYEASCNLFS 600

QY 600 KDPYSILYQKNIEGSETAYYYYADAETKEIDKYRIPYOISNKNIKLTFIGHGKSEFN 659
DB 601 KNPYDILYQKNIEGSETAYYYYADAETKEIDKYRIPYOISNKNIKLTFIGHGKSEFN 660

QY 660 TDTFANLVDVLSSEIETTLNKAADISPKYTEINLLGONMFSYSABEYTPGKLLKI 719
DB 661 TDTFANLVDVLSSEIETTLNKAADISPKYTEINLLGONMFSYSABEYTPGKLLKI 720

QY 720 KDRVSELMPSIQSDSITVSANOVEVRINEEGKEIILDHSGKWNKEESIIKOISSKEYIS 779
DB 721 KDRVSELMPSIQSDSITVSANOVEVRINEEGKEIILDHSGKWNKEESIIKOISSKEYIS 780

QY 780 FNPENKIIIVSKYKYLHLSLTLQOIRNANSSDIDLEKKYMLTECEINVASNIDROAVEG 839
DB 781 FNPENKIIIVSKYKYLHLSLTLQOIRNANSSDIDLEKKYMLTECEINVASNIDROAVEG 840

QY 840 RIEBAKNTSDSINYIKNEFKLIESISDSLYDLKHQGLDDSHFISFEDISKTEGFRIR 899
DB 841 RIEBAKNTSDSINYIKNEFKLIESISDSLYDLKHQGLDDSHFISFEDISKTEGFRIR 900

QY 900 FINKETGNSIFETEKEIFSEYATHISKEISNIDKTIIDNVNGKLVKKVNLDAAEHVNTL 959
DB 901 FINKETGNSIFETEKEIFSEYATHISKEISNIDKTIIDNVNGKLVKKVNLDAAEHVNTL 960

QY 960 NSAFFIQSLIEYNTKESLSNLSVAMQVYVAQLFSTGLNTITDASKKVELVSTALDETI 1019
DB 961 NSAFFIQSLIEYNTKESLSNLSVAMQVYVAQLFSTGLNTITDASKKVELVSTALDETI 1020

QY 1020 DLLPTLSEGLPIIATIDGVSIGRAIKELSETNDPLLRQEIKAIGIMAVNLTAASIAIV 1079
DB 1021 DLLPTLSEGLPIIATIDGVSIGRAIKELSETNDPLLRQEIKAIGIMAVNLTAASIAIV 1080

QY 1080 TSALGIASGFSILLVPLAGISAGIPSLVNNELIILQDKATKVIDYFKHISLAETEGFTLL 1139
DB 1081 TSALGIASGFSILLVPLAGISAGIPSLVNNELIILQDKATKVIDYFKHISLAETEGFTLL 1140

QY 1140 DDKIIMPDDVLVLSIDFNNNISITLKCCEIWRAGEGSGHTLTDIDHPFSPSTYRKPM 1199
DB 1141 DDKIIMPDDVLVLSIDFNNNISITLKCCEIWRAGEGSGHTLTDIDHPFSPSTYRKPM 1200

QY 1200 LSIYDVNLKKEKIDFSDKOLMVLNAPNRVFGYENGWTPGPRSLDNDGTLLDRIDRHYE 1259
DB 1201 LSIYDVNLKKEKIDFSDKOLMVLNAPNRVFGYENGWTPGPRSLDNDGTLLDRIDRHYE 1260

QY 1260 GQFYWEYFAFIADALITLKPVEYEDTVRINLDGTRSFIVPVIITEOIRKNLSYSFYGS 1319
DB 1261 GQFYWEYFAFIADALITLKPVEYEDTVRINLDGTRSFIVPVIITEOIRKNLSYSFYGS 1320

QY 1320 GGSYSLSLSPYNNMTDLNLVENDTVIDVNVNKNITTESDEIOKGLIENILSKNITED 1379
DB 1321 GGSYSLSLSPYNNMTDLNLVENDTVIDVNVNKNITTESDEIOKGLIENILSKNITED 1380

QY 1380 NKIILNNHTINFGYNESNRFISLTFSEIDENIIEIDLVSKSYKILLSGNCKMLIEN 1439
DB 1381 NKIILNNHTINFGYNESNRFISLTFSEIDENIIEIDLVSKSYKILLSGNCKMLIEN 1440

QY 1440 SSDIOQKIDHGFNGEHQKYPYSDIDNETKNGFIDYKKEGLPTAEFSNESIIRNTYM 1499
DB 1441 SNHIOQKIDHGFNGEHQKYPYSDIDNETKNGFIDYKKEGLPTAEFSNESIIRNTYM 1500

QY 1500 PDSNNLFIYSSKDLKDIRIINKGDVLLIKNYFKDDMKVLSFTIEDTNTIKNGVYLDE 1559
DB 1501 PDSNNLFIYSSKDLKDIRIINKGDVLLIKNYFKDDMKVLSFTIEDTNTIKNGVYLDE 1560

QY 1560 NGVAQILKEMNNAKALANTSLSMNFLESINIKNIFNNLDPNIEFIIDTNPFIISGNSI 1619
DB 1561 NGVAQILKEMNNAKALANTSLSMNFLESINIKNIFNNLDPNIEFIIDTNPFIISGNSI 1619

QY 1620 GQPELLCDKDKNIQPYFINFKIKTSYTLVGNRQNLIVEPSYHLLDDSGNLSSTVINFSQ 1679
DB 1620 GQPELLCDKDKNIQPYFINFKIKTSYTLVGNRQNLIVEPSYHLLDDSGNLSSTVINFSQ 1679

QY 1680 KYLYGIDRVYKVIAPNLTYT 1700
DB 1680 KYLYGIDRVYKVIAPNLTYT 1700

RESULT 4
Q46034
ID Q46034
AC Q46034;

PRELIMINARY; PRT; 2367 AA.

[illegible]

Qy	540	EGALGEDDNDLDPAAQ	553	
Db	541	EGALGEDDNDLDFSQ	554	
RESULT 6				
Q46149		PRELIMINARY;	PRT;	2178 AA.
ID	Q46149	Q46147; Q46148;		
AC	Q46149; Q46147; Q46148;	01, Created)		
DT	01-NOV-1996	(TREMELrel. 01, Last sequence update)		
DT	01-NOV-1996	(TREMELrel. 01, Last sequence update)		
DT	01-JUN-2003	(TREMELrel. 24, Last annotation update)		
DE	Alpha-toxin.			
OS	Clostridium novyi.			
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;			
OC	Clostridium.			
NCBI	NCBI_TaxID=1542;			
[1]				
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC19402;			
RC	MEDLINE=95342160; PubMed=7616958;			
RA	Hofmann F., Herrmann A., Habermann E., von Eichel-Streiber C.;			
RT	"Sequencing and analysis of the gene encoding the alpha-toxin of			
RT	Clostridium novyi proves its homology to toxins A and B of Clostridium			
RT	difficile."			
RL	Mol. Gen. Genet. 247:670-679 (1995).			
RN	[2]			
RP	SEQUENCE OF 1204-2178 FROM N.A.			
RC	STRAIN=ATCC19402;			
RA	Hofmann F., Habermann E., von Eichel-Streiber C.;			
RL	Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; Z48636; CAA88565.1; -			
DR	EMBL; Z23280; CAA80818.1; -			
DR	EMBL; Z23281; CAA80819.1; -			
DR	PIR; S55805; S55805.			
DR	InterPro; IPR002479; CW binding.			
DR	InterPro; IPR007577; Gly transf sug.			
DR	Pfam; PF01473; CW binding 1; 10-			
DR	Pfam; PF04488; Gly transf_sug; 1.			
DR	SEQUENCE 2178 AA; 250134 MW; 980ADCE031C4A75A CRC64;			
Qy	3	LVNKAQLQKVVYKFRIOEDYVAILNALEEH-NMSESVVEKYLKLDINLLTDNYLN	61	
Db	2	LITREQMLKTASIPLRKKEPEYNIILDALENFRDIEGTSVKIYKSLKCNELVDNYQT	61	
Qy	62	TYKSGRNKALKFKEVLTWVEVLKONSLTP--VEKNLHFITGGQINDTAINYINQWK	119	
Db	62	KYPSGGRNLALENFRSLYSELRELTKNSTSTIAKSLFIWGGPISQSLYYNNWK	121	
Qy	120	DVNSDYTVKYFDSNAFLINTLTKTIVESTNTNLSFRLENLNDPEFDYKFKRKEI	179	
Db	122	MFNKDYNIRLFYDKNSLLVNTLTAIOESSKVIIEQNSNILDGTGHNKFYSDRMKLI	181	
Qy	180	YDQKHFIDYKSGIENPFIDNIKTLYSNEYSKDLALNKYIEESLNKLTANNND	239	
Db	182	YRYKRELKMLYENKQKNS--VDDIIFNLSFYKDYDIGKANQKNNKKMAIGATD	238	
Qy	240	IRNLEKFAEDLVRLYNOELVERWNAAAADIIRLSMKEDGGVYLLDVIDLPQTDLFK	299	
Db	239	I-NTENILTNKLKSYVYQELIQTNNLAAAADILRIALKKVGVCDDLFLPGVNLSEFN	297	
Qy	300	SINKPDSITNTSWEMIKLEAIMKYKEVPGYTSKNFMDLDEEVSQSFESALSSKDKSEI	359	
Db	298	DISKPGMDSNYEAAIFEAIAENEKLMNNYPKYMEQVPSSEIKERILTSFVRN-HDINDL	356	
Qy	360	FLPLDDTKVSELEV-----KIAPANNSVINQALISLSDYSQDLVINQIKNRYKILNDNL	414	
Db	357	ILPLGDIKIQSEILLRLKAATGKTKFSNAFTIISNDSITLNNLISOLENRYEILASII	416	


```
QY 415 NP--SINEGTD--FNTTMIKIFDGLASISNEDNMFMKIITNYLVKGFAPDVRSTINLSG 470
Db 417 QEKFKICEYDYSVSVSELVLETPKNSLMDGSSPYQOICGLSSGFKPEVNSVFFSG 476
QY 471 PGVYTAYQDLMLFXDN-----STNIHLLEPELNFEPPTKISQLTEQBITTSUWSNQ 524
Db 477 PNIYSATCDTYHFIKNTFMDLSSQOEIPEAS--NNLYFSK-----THDEFKSSWLLRS 529
QY 525 ARAKSOFEYKGYFGALGEDNDLDPQNTVLDKDYVSKKI--LSSMKTRNKEVHIHV 582
Db 530 NIAEKEFOKLITYIGRTLYNEDGLNFKNKRVTTSELKVBEEVNSTKIYENYDLNMIL 589
QY 583 QLOGDXISVEASCNLFSPKDPYSILYQKNIEG--SETAYVYVYVADAEIKEDIKRYPYOIS 641
Db 590 QIQGDDISVESAVNVFGKPNKAIL-----IQGVDDFANVYFENGIVQSDNNILNLSRN 645
QY 642 NKENIKLITIGHGKSEFNTDTFANLVDLSLSBEI-----ETILNLAADISPKYIEINLL 696
Db 646 DIKKIKLTLIGHGENVFNPKLFGKTVNDLYTNIKKPKLOHLLEREGVILKNKYLKINIL 705
QY 697 GCNMFYSYSABETYPGKLLKIKORVSELMP--SISQDSITVSANOYEVRIINEBKEIL 755
Db 706 GCMYFTPKVDINGTFVGLKFNKIS---RDLQPKGFSGKNQLEISANKYAIRINREGKRYL 762
QY 756 DHSGKWINKESIIKDISKEYISPNKENKIIIVKSKYLHELSTLIQETIRNNANSSDIDL 815
Db 763 DYFGKWSNTDLIAEQISNKYVYVNEVENTLSARVEQLNKVAFKAI-----NSIIQT 817
QY 816 EKKVMLTECEINVASINDROIVEGRIEEAQNLTSDSI---NVKNEFKLIESISDLYD 871
Db 818 TNNQELKQSLVNYADLITLYSELKEDIPPELDNIQIKERILNELISLHDFSNILID 877
QY 872 LRHQGLDDSHFISFEDISKTENGFRIRFINKETGNSIFIETBEKEIFSEYATHISKEIN 931
Db 878 FYQKNNISNMIIIFDSIIKEKYVNVKLANKITGETSVIKTYSDSLWNFTNKYKIVDD 937
QY 932 IKOTIFPNVNGKLVKNLDAAEHVNTLNSAFIQLSIIENYTKESLNSLVAMKQVYA 991
Db 938 IKGIIVKDIINGEFIKADFEIQNPSSLNAMLQULLIDYKPYTEILTNMTSLKQVYA 997
QY 992 QLFSTGLNTITDASKVVELSTALDEITDILPTLSGLPIIATIIDGVSIGAAIKELSET 1051
Db 998 QIFOLSIGAQEATEIVTIIISDALNANFNILSKLVGSSVASVIIDGINLIAALTELKNV 1057
QY 1052 NDPLLROEIEBAKIGIMAVNLTAATAIVTSAIG---IASQFSILLVPLAGISAGISLVN 1108
Db 1058 KTNFERKLEAKVGMYSIGIFLSSLSIGLGTAVSELIGVSVPGILVGLSLAVN 1117
QY 1109 NELILODKATKVIDYFKHISLAETEGFTLLDDKIIIMPQDVLSEIDFNNSNITIGKE 1168
Db 1118 NILVLGEKYNQILDYFSKFPYIVGKNPFS-IQDNIIPYDDIAITELNFKYKFKYGYAK 1176
QY 1169 IWRAGSGGHTLTDIDHFPSSPSITVRKPLWSIYDVLTNKKKIDPSKDLVLPNAPNR 1228
Db 1177 ISGLKVLGVLTIGHENIDHYFAPSLDHYIE--LSIYPALKLNDTNLP--KGNVLLPSGLNK 1234
QY 1229 VFGYEMGWTPGFRSLDNDGKLLDRIRDHV-----EGCFWRYRFAFTADALITKLAPRYE 1283
Db 1235 VYKPEISAIAGANSQEGNGVEVLNLIIRNYVDSNGNTKFPKYEAPP--EYSFYSRMEYEF 1293
QY 1284 DTNVRILNDGNTRGFIVPVITTEQIRKNLSYFSGSGSYLSLSPYNNMIDMLVENDT 1343
Db 1294 DTKVNVILDNENKTLIIPVLITIDEMRNKISYEILGDGQYNVILPVNQTNIVSNKNDI 1353
QY 1344 WVIDVNVKNITTESDIEQKELIENILSKLNTIEDNKIILNHTNIFYGDINESRPFIS 1403
Db 1354 WNFVDSYIVKESKIEDNKFLVDGFINNFISTLKVSDNGFKIGQFIS---IKNTPRAIN 1409
QY 1404 LTFSLIEDNIIILIDIL--VSKSYKILSCNMKLIENSDD-IQQKIDHIGFNGEHQYIP 1461
Db 1410 LSFKINNNI--VIVSYTLNHEKSNSTIITSSDLNDIKNNFNLDNINNYIGLSISDNTIN 1468
QY 1462 YSIDNETKYNGFDIDYSKKEGLFTAEPGNEBESIIIRNIYMPDNNILFIYSSKDLDIRINK 1521
```

```
Db 1469 CIVRNDEVYMEGKI-----FLNEK--KLVFIOQNELEHLHYDS-----VNK 1506
QY 1522 GDVKLLIGN-----YFKDDMKVSLGSTIEDT-----NTIKLVGVLDENGVAQI 1565
Db 1507 -DSOYLNNFINNVVKOGYIVVEGFLINSTENKYSLYIENNKIMLKGILY-ESSVFKT 1564
QY 1566 LKFMNNAKSALNT--SNSLMNFLSINIK---NIFYNNLDPN---IFILDTN--FIISGS 1616
Db 1565 IQDKIYSKEVNDYILSLIKKFTVNIQLCFPMVSGVDNENRYLEVMLSTNNKWIING- 1623
QY 1617 NSIQOFELICDKNIOPIYFINPKIKETSITLVGNRQNLIVEPSYHLDSDGNISSITVIN 1676
Db 1624 --GYWE---ND-----FNNYKIVDF-----EKCNIIVSGSNKLNSEGLADT-ID 1662
QY 1677 PSQKLYCIDRYNKNVIAAPLXYT 1700
Db 1663 VLDKOLENL--YIDSVIIIPKXYT 1684

RESULT 7
O88653 PRELIMINARY; PRT; 698 AA.
AC O68653;
ID O68653;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Truncated toxin A.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=CCUG 20309;
RX MEDLINE=99304497; PubMed=10376209;
RA Song K.P., Bai X.L., Chang S.Y.;
RT "Nucleotide and peptide sequences of the open reading frame encoding a
RT truncated toxin A gene of Clostridium difficile strain CCUG 20309.";
RL DNA Seq. 10:93-96(1999).
DR EMBL; AF053400; AAC08437.1; -.
DR InterPro; IPR007577; Gly_transf_sug.
DR Pfam; PF04488; Gly_transf_sug, I.
SQ SEQUENCE 698 AA; 80682 MW; 021EB268A3BDECSE CRC64;

Query Match 22.1%; Score 1916; DB 2; Length 698;
Best Local Similarity 52.9%; Pred. No. 3.8e-65;
Matches 371; Conservative 138; Mismatches 184; Indels 8; Gaps 5;

QY 1 MNLVNAQLQKVVYKFIQDEYVAIUNALEEYHNMSSESVVEKYLKLDINLTDNYL 60
Db 1 MSLISKBELIKLAY-SIRPRENEVKTVLNLDYENKLTNNENKYLQKKLINESIDVFM 59
QY 61 NTYKSGRNKALKKFKVELTWEVLELKNLSLTPVEKNLHFTWIGQINDTAINYNQWD 120
Db 60 NKYKSSNRNALSNLKDILKEVILIKNSNTPVEKNLHFWIGGVEDIVLEYIKQWAD 119
QY 121 VNSDVTYKVFYDSNAFLINTLKKTIVESATNNTLESFRENLDNDFYKFKREMEIY 180
Db 120 INAETNIKLVYDSEAFVNTLTKAIVESTTEALQLLEEIQNCFQDNMKYKKEMEFIY 179
QY 181 DKQKHFDIYKYSQIBENPEFIIDNIIRKTYLSNYSKOLEALNKYTEESLNKITANGNDI 240
Db 180 DRQKRFINYKSIQINKPTVLTIDDIKSYLVSEYNRDETLLSYRTNSLRKINSNHDIV 239
QY 241 RNLEKFAEDLVRNLQBELVERWNLAASDILRISMLKEDGGVYLDVILPGIQDLPKFS 300
Db 240 RANSLFTEOELLNIYNQELLNRGNLAAASDIVRLALKNFGVYLDVMDLPQIHSDLPKT 299
QY 301 INKPSITNTSHEMIKLEIAIMKYKEVIGYTSKNFMDLDEEVQSFESALSSKSKSEIF 360
Db 300 ISRPSSIGLDRWEMIKLEIAIMKYKYYINNYTSNFDKLDQDQKQNFKLIIESKSEIF 359
```



```
QY 361 LPLDDIKVSPLEVKIAPANNVINOALISLKDSCVSLVINOIKRKYKILNDNLNPSINE 420
Db 360 SKLENLNVSLDEIKIAPALGVSINOALISKSGSVLNLVIEQIKRKYQFNLQHLNPAIES 419
QY 421 GTDFTNTMKIPSDKLASISNEDNMFMKITNYLKVGFAPDVRSTINLSGPGVYTGAYQD 480
Db 420 DNNFTDTTKIPHDSLFNSATAENSMTLTKIAPYLQVGFMPPEARSTISLSGPGAYASAYD 479
QY 481 LLMFKDNSTNIHLLEPELRNPFEPKTKISQTEQBITSLMSPFNQARAKSQFEYKKGYPF 540
Db 480 FINLQENTIEKTLKASDLIEFKFPENNLSQTEQBITSLMSPFNQARAKSQFEYKKGYPF 539
QY 541 GALGEDDNLDFPAQNTVLDKDY-VSKKILSS--MKTRNKEYTHYIVQLQGDKIYEASCNL 597
Db 540 GSLSEDNEVDNFKNTALDKNYLLANKIPSNVVEAGSKNYHYIIQLQGDIISEATCNL 599
QY 598 FSKDPYSSILYQKNIEGSETAYYYVAD--AEIKEDIKYRIPYQISNKRNIKLTFIGHGK 655
Db 597 FSKDPYSSILYQKNIEGSETAYYYVAD--AEIKEDIKYRIPYQISNKRNIKLTFIGHGK 657
QY 656 SEFNTDTPANLVDVSLSEIEITILNLAKADISPKYIEINLL 696
Db 658 DEFNTSEFARLSVDSLSNEISSFLDTIKLIDISPKNVEVNL 698
RESULT 8
Q9EXQ8 PRELIMINARY; PRT; 698 AA.
AC Q9EXQ8;
DT 01-WAR-2001 (TREMELrel. 16, Created)
DT 01-WAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Truncated toxin A.
GN TCDA.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8864;
RA von Eichel-Streiber C.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=8864;
RA Kohl M.;
RT "The pathogenicity locus of Clostridium difficile, strain 8864.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011301; CAC19893.1; -.
DR InterPro; IPR007577; Gly_transf_sug.
DR Pfam; PF04488; Gly_transf_sug; I.
SQ SEQUENCE 698 AA; 80799 MW; 66D9B21798D3149B CRC64;

Query Match 21.9%; Score 1898; DB 2; Length 698;
Best Local Similarity 52.4%; Pred. No. 1.8e-64;
Matches 367; Conservative 140; Mismatches 186; Indels 8; Gaps 5;

QY 1 MNLVNAQLOKQVYKFIQDEYVAIINALSEYHNMSESSVVEKYLKLDKINLLDNYL 60
Db 1 MSLISKEELIKLAY-SIRPRENEYKTVLTNLDEYNKLTNNENKYLQKKLINESIDVFM 59

QY 61 NNYKSGRNKALKKPKYLTMTVEIKNLSLTPVEKNLHFTWIGQINDTAINYNQW 120
Db 60 NKYKSSRNALSNLKKOLKEVILKNSNTSPVEKNLHFTWIGGEVSDIVLEYIKQW 119

QY 121 VASDYKVKYVFNAPLINTLAKTIVESATNTNLTSEFNLNDPFDYKFKEMELIY 180
Db 120 INAEYNIRLWYSEAFNLNLKKAIVESSTTALQILSEEIQNPQFDNKKYKEMELIY 179

QY 181 DKQKHFIDYKQIENPEFIIDNIKTYSNEYSKOLEALNKYIEESINKITANGNDI 240
Db 180 DRQKRFINYKXSQINKPTVLTIDDIKSYLISEYNRDETLLSEYRTNSLRKNSNHGIDV 239
```

```
QY 241 RNLEKFADEDLRLVNOELVERWNLAASADILRISMLKEDGGVYLDVDILPGIQDPLFS 300
Db 240 RANSLFTEQELNLYNQELLNARGLAASADIVRLALKNFGVYLDVDMPLGIHSDLFKT 299
QY 301 INKPDSTINTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQSFSAUSSKSKSEIF 360
Db 300 ISRPSSIGLDRWEMIKLEAIMKYKYINNYTSENFDKLDQOLKDNFKLIIIESKSEKCEIF 359
QY 361 LPLDDIKVSPLEVKIAPANNVINOALISLKDSCVSLVINOIKRKYKILNDNLNPSINE 420
Db 360 SKLENLNVSLDEIKIAPALGVSINOALISKSGSVLNLVIEQIKRKYQFNLQHLNPAIES 419
QY 421 GTDFTNTMKIPSDKLASISNEDNMFMKITNYLKVGFAPDVRSTINLSGPGVYTGAYQD 480
Db 420 DNNFTDTTKIPHDSLFNSATAENSMTLTKIAPYLQVGFMPPEARSTISLSGPGAYASAYD 479
QY 481 LLMFKDNSTNIHLLEPELRNPFEPKTKISQTEQBITSLMSPFNQARAKSQFEYKKGYPF 540
Db 480 FINLQENTIEKTLKASDLIEFKFPENNLSQTEQBITSLMSPFNQARAKSQFEYKKGYPF 539
QY 541 GALGEDDNLDFPAQNTVLDKDY-VSKKILSS--MKTRNKEYTHYIVQLQGDKIYEASCNL 597
Db 540 GSLSEDNEVDNFKNTALDKNYLLANKIPSNVVEAGSKNYHYIIQLQGDIISEATCNL 599
QY 598 FSKDPYSSILYQKNIEGSETAYYYVAD--AEIKEDIKYRIPYQISNKRNIKLTFIGHGK 655
Db 600 FSKNPKNSIIQORNV--NESAKSHFLSDGSDIELNKYRIPERLNKKEKVKVTFIGHGK 657
QY 656 SEFNTDTPANLVDVSLSEIEITILNLAKADISPKYIEINLL 696
Db 658 DEFNTSEFARLSVDSLSNEISSFLDTIKLIDISPKNVEVNL 698
RESULT 9
O86141 PRELIMINARY; PRT; 697 AA.
AC O86141;
DT 01-NOV-1998 (TREMELrel. 08, Created)
DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE TcdA protein (Fragment).
GN TCDA.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8864;
RA Schein F.; Leukel P.; Weidmann M.; Eichel-Streiber C.V.; Braun V.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=8864;
RA von Eichel-Streiber C.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y10689; CAA71690.1; -.
DR InterPro; IPR007577; Gly_transf_sug.
DR Pfam; PF04488; Gly_transf_sug; I.
FT NON TER
SQ SEQUENCE 697 AA; 80680 MW; CD650D3372D65B73 CRC64;

Query Match 21.8%; Score 1892; DB 2; Length 697;
Best Local Similarity 52.4%; Pred. No. 3.1e-64;
Matches 366; Conservative 139; Mismatches 186; Indels 8; Gaps 5;

QY 3 LVNAQLOKQVYKFIQDEYVAIINALSEYHNMSESSVVEKYLKLDKINLLDNYLNT 62
Db 2 LISKEELIKLAY-SIRPRENEYKTVLTNLDEYNKLTNNENKYLQKKLINESIDVFM 60

QY 63 YKSGRNKALKKPKYLTMTVEIKNLSLTPVEKNLHFTWIGQINDTAINYNQW 122
```

Db 61 YKSSRNRLSNLKDILKEVLKNSNTSPVEKNLHFVWIGVESDIVLEVIKQWADIN 120
Qy 123 SDYTVKVFDSNAFLINTLKTIVESATNTLESFRENLDNBEFDYKRYKRMELIYDK 182
Db 121 AYNIRLWYDSEAFVNTLKKAIVESSTTEALQLEEBEQNFQFDMKFKYKRMFIYDR 180
Qy 183 QKHFIDYKSOIEENFEFIDNIITKYLSEYNSKDLKALNKYIESLNKITANNNDIRN 242
Db 181 QKHFIDYKSOINKPVTITDIIKSYLSSEYNRDTELESRTSLRKINSHGIDVRA 240
Qy 243 LEKFADEDLRLYNQBELVERNWLAAASDILRISMLKEDGGVYLDVDILPGIDPLFKSIN 302
Db 241 NSLFTQEGLLNTYQELLNARGNLAASDIVRLALKNFGVYLDVDMFGIHSDFLKITS 300
Qy 303 KPDSITNTSWEMIKLEAIMKYKYPGYTSKNFMDLDEEVQSPFSALESKSKSEIFLUP 362
Db 301 RPSSIGLDRWEMIKLEAIMKYKYNNTSENPKDQDKONFKLIIESKSEKSEIFSK 360
Qy 363 LDIKVSPLVKIAPFANNSVINOALISLSDYSCSDIVINQIKRYKILNDNLNPSIN 422
Db 361 LENLNSDLEIKIRFALGVSINOALISKSGSYLTNIVIQKRYOFLNQHLPAPAESDN 420
Qy 423 DFNFTWKIFSDKLASINEDNMFMKITNYLKVGFAPDVRSTINLSGGVYTGAYQDILL 482
Db 421 NFDYTKIFHDSLFNSATAENSFUTKIAPYLOVGFMPPEARSTISLSGGAYASAYYDFI 480
Qy 483 MPKDNSTNIHLPELRNFPFKTKISQTEQBITSLWFSNQARAKSQPEEYKKGFEGA 542
Db 481 NLQENTIEKTLKASDLIEFKFPENNLSQTEQINSLSWFSNQARAKSQPEEYKKGFE 540
Qy 543 LGEDNDLPAQNTVLDKDY-VSKILSS--MKTRNKEYIHYIYVLOGDKISYEASCNLFS 599
Db 541 LSENEVDENKNTALDKYLLNNKIPSNVVEEAGSNVYHYITLQGGDDISYEATCNLFS 600
Qy 600 KDPYSSLYKNTGSETAYVYVAD--ABIKEIDKVRIPYQISNKRNTKLFIHGKSE 657
Db 601 KNPKNSTIIQRM--NESAKSHFLSDGSEILSNKYRIPERLKNKEXKVYFIHGKDE 658
Qy 658 FNTDTPANLDVSLSSBIETILAKADISPKYIEINLL 696
Db 659 FNTSEFARLSVDSLSNEISSFLDTIKLIDISPKNVEVNLL 697

RESULT 10
Q9FCX5 PRELIMINARY; PRT; 553 AA.
AC Q9FCX5; 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Toxin A (Fragment).
GN TCDA.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OC NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C34;
RX MEDLINE=20402122; PubMed=10931294;
RA Braun V., Mehlig M., Moos M., Rupnik M., Kalt B., Mahony D.,
RA von Eichel-Streiber C.;
RT "A chimeric ribozyme in Clostridium difficile combines features of
RT group I introns and insertion elements."
RL Mol. Microbiol. 36:1447-1459(2000).
DR EMBL; AJ131844; CAC03681.1; -.
DR InterPro; IPR007577; Gly transf. sug.
DR Pfam; PF04488; Gly transf. sug; 1.
FT NON-TER 553
SQ SEQUENCE 553 AA; 63913 MW; DDE4551A6D8C3B25 CRC64;
Query Match 17.0%; Score 1474; DB 2; Length 553;
Best Local Similarity 51.7%; Pred. No. 1.7e-48;

Matches 287; Conservative 108; Mismatches 158; Indels 2; Gaps 2;
Qy 1 MNLVNAQLOKQVYKEROEDEYVAILNALAEYHNMSSESVVEKYLUKLDINNLTNYL 60
Db 1 MSSIKKELIKLAY-SYRPRENEYKTLTWLDEYNKNTNNENKYLOLKUESIDVFM 59
Qy 61 NTYKSGRNKALKKFKELYLTWEVLELNKNSLUTPEVKLHFITWGGQINDTAINYQWMD 120
Db 60 NKYNSSRNRLSNLKDILKEVLKNSNTSPVEKNLHFVWIGVESDIALEVIKQWAD 119
Qy 121 VNSDYTVKVFDSNAFLINTLKTIVESATNTLESFRENLDNBEFDYKRYKRMELIY 180
Db 120 INAEYVNLWYDSEAFVNTLKKAIVESSTTEALQLEEBEQNFQFDM-KFYKRMFIY 178
Qy 181 DKQHFIDYKSOIEENFEFIDNIITKYLSEYNSKDLKALNKYIESLNKITANNNDI 240
Db 179 DRQRFINYKSOINKPVTITDIIKSHLVSEYNRDETLESYRTSLRKINSHGIDI 238
Qy 241 RNLEKFADEDLRLYNQBELVERNWLAAASDILRISMLKEDGGVYLDVDILPGIDPLFKS 300
Db 239 RANSLFTQEGLLNTYQELLNARGNLAASDIVRLALKNFGVYLDVDMFGIHSDFLKT 298
Qy 301 INKPDSTNTSWEMIKLEAIMKYKYPGYTSKNFMDLDEEVQSPFSALESKSKSEIF 360
Db 299 IPRSSIGLDRWEMIKLEAIMKYKYNNTSENPKDQDKONFKLIIESKSEKSEIF 358
Qy 361 LPLDDIKVSPLEVKIAPFANNSVINOALISLSDYSCSDIVINQIKRYKILNDNLNPSINE 420
Db 359 SKLENLNSDLEIKIRFALGVSINOALISKSGSYLTNIVIQKRYOFLNQHLPAPAES 418
Qy 421 GTDFTWKIFSDKLASINEDNMFMKITNYLKVGFAPDVRSTINLSGGVYTGAYQD 480
Db 419 DNFDTTKIFHDSLFNSATAENSFUTKIAPYLOVGFMPPEARSTISLSGGAYASAYYD 478
Qy 481 LLMFKDNSTNIHLPELRNFPFKTKISQTEQBITSLWFSNQARAKSQPEEYKKGFE 540
Db 479 FINQENTIEKTLKASDLIEFKFPENNLSQTEQINSLSWFSNQARAKSQPEEYKKGFE 538
Qy 541 GALGEDNDLPAQNT 555
Db 539 GSPSGDNGVDENKNT 553
RESULT 11
Q9ZGR4 PRELIMINARY; PRT; 3169 AA.
AC Q9ZGR4;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative cytotoxin.
GN L7095.
OS Escherichia coli O157:H7.
OG Plasmid pO157.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OC NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EDU933;
RX MEDLINE=98391744; PubMed=9722640;
RA Burland V., Shao Y., Perna N.T., Plunkett G., Sofia H.J.,
RA Blattner F.R.;
RT "The complete DNA sequence and analysis of the large virulence plasmid
RT of Escherichia coli O157:H7."
RL Nucleic Acids Res. 26:4196-4204(1998).
DR EMBL; AF074613; AAC70163.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR InterPro; IPR007577; Gly transf. sug.
DR InterPro; IPR006473; Peptidase_C38_yt.
DR Pfam; PF04488; Gly transf. sug; 1.
DR TIGRFAMs; TIGR01586; yopT_cys_prot; 1.
KW Plasmid.

SQ SEQUENCE 3169 AA; 361971 MW; C2CDB48A72BD5FD1 CRC64;

Query Match 5.9%; Score 512.5; DB 2; Length 3169;
Best Local Similarity 20.2%; Pred. No. 4.6e-11;
Matches 422; Conservative 333; Mismatches 720; Indels 619; Gaps 104;

QY 14 YVKEFI--QEDVAVIALNAL--EYHNMSSESVVYKYLKDDINLTNYLNTYKKSGR 68
DB 160 YIKIRKTRGABDQTTTQSLIINELLNGVDRNTI-PFQKISELNDIHSYENQJKNR 218
QY 69 N--KALKKFEYLWVLELKNS--LTPVERKLFHFWIGGQ 106
DB 219 KGIILYKQGLSSLNDKNGKQSLDNASKIINILGIEVQSHKVDIEFFIHAVVAGA 278
QY 107 INDTAINYINQWQNSDYTVKPYDSNAF--LINTLKT-- 144
DB 279 PPDNTFSYITAFNTYKDYTYLLWIDPNAFAAKFSGILKNIAMYALMELRRTNPHLAE 338
QY 145 -----IVESATNTLE--SPRENANDPEFY-----NKEYRKMELIYDKQKHFI 197
DB 339 EMNEVILKIQIONETIEFKETRELRKLELNRYSKLTSETKEKPNVFFLESMTGMQDNYF 398
QY 188 DYKSOIBENPEFT-----IDNLIK--TYSNEYSKDEALNKYIEBSLNKIITANNGN-- 238
DB 399 TYCISNGISNTDDISRLDFLNVKLSPFVONDPKSTVEKNKRDIDILLKNTISQKGDPR 458
QY 239 ---DIRNLEKFADELVLYNQLVERNLAAASDIILRMLKEDGGVYLDVILPGIQP 295
DB 459 QLRDINTLESFKKQDFFYQOEMLRWNYAAASDQVRINILKEYGGIYTDILPAYSD 518
QY 296 DLFSKINKPDSITWSEMIKLEIMKY--EYTPG--YTSKNPMDLDEEVQSFSAIS 351
DB 519 KVSQIINE-KSDKRRFFEDLRLRIISGISILSKGKYSIKH-DGLDETTLAQLNAIL- 575
QY 352 SKSDKSEFLDLDIKVSPLEVKIAFANNVI-----NQALISLKDSYC 395
DB 576 SEIEK-----LTDYD-FKPVETKVVRDTFKFYKQYKWTWNIRNNFMTHKSKC 630
QY 396 SDLVINQIKRYKILNDNLNINSGTDFNTMTKIFSDKLASINE-----DNM 444
DB 631 IDFILSGQKQYLEL-QRIRNISIYNNLFYTT-----EDLSLNVAIGGIPAKKYLEHG 684
QY 445 MFMIKITNLYKVGAPDVRSTINSGP----- 471
DB 685 LF---SEYRQDGTIPYVSTLNSGPDIMQMKYKSLGRIGEVHAIKONKLSVDNPL 740
QY 472 GYTCAYQDLMF---KDNSTNIHLPE-----LNFEPKTKISQLT-----EQBIS 518
DB 741 GYASNSKDNKSFWNLNPSVSGINDITPDDSSWAVRNNDINKILFEKINCHVEKLPTS 800
QY 519 LWSFNQARAKSQFEYKKGYPGALGEDDNLDFACNTVLDKDYVSKILSMKTRNKEYI 578
DB 801 LY-----YEIDRSFFQW---DNKSIRHVTIEINKDLI--KQINLLTSSNIDV 844
QY 579 HYIVOLQDKISYEASCNLFSDPYSSILYQKNIEGSETAYVYVADAEIKEDIKRIPY 638
DB 845 KLLIKL--DRELYAISSKI-----DNPALARSITLQLQANVTSNTPEPNTINFIY 896
QY 639 QISNKN-----IKLTTIGHKSEFNTDTFANLDDVS-----LSSEIET 677
DB 897 DFKKKQDLLSAIKL-----FSRNDADTKIIWYNSVMKRVFLREVISC 942
QY 678 IINLAKADISPKYIENLLGNMPSYSISAEETPGKLLIKIKORVSELMPSISQDSI-- 735
DB 943 VLRSKKVD---SYINEN-----KXNLSKEDA--GALRDYAKUKOKELFNLDDDDGYKK 990
QY 736 TVSANQYEVRIINEEGREILDSKWKINKESIIKDISKEYISNP-----K 783
DB 991 IITWAY---IKERKL-----SGIYVNIENSIISGHESFDIIRSNQHWGDLSTVBQFK 1042
QY 784 ENKIIVSKYLHELSTLQEIERNANSDDIDLEKKVMLTECEINVAS-----NIDRQIVE 838
DB 1043 KFEFYKSE-LSSAKSIFDDIKNK-VITDPEKRVNLYHQLDSDIKERIAFLDISHVAYP 1100

QY 839 GRIBEAKNLSDSYINIKNEFKLIESI-----SDSLYDLKHQNL 878
DB 1101 GSLLEKLOLS---GYFSDINIIEAVLLASYGVSHGVVYVAPSDKLELLRHTK 1156
QY 879 DSHFISPEDISKTEGNGFRIRINKETGNSI---FIETEKIEFSEYATHIKESINIKDT 935
DB 1157 SNSEWI--EKITP---YVYDILSDNSVNLRPPLSEBQKILNDIKLEISKSVS--EQ 1207
QY 936 IPDNVNGKLVKKNLDAAEHVNTLNSAPFIOSLIYNTTK-----ESLSNLSV----- 983
DB 1208 YFMKLTEQSSVIGIKYSVDFRYNENFLSLPQNQLTLPFMYRYFEMLYDIHIGIEN 1267
QY 984 AMKVQVVAQLFSGLNTITDASKVVELV-----STAL-DEITDL 1021
DB 1268 KANREFTYKSSUNLDFLINDERVNLGLIKYKYLSEIHRILTNSISPADISPL 1327
QY 1022 LPTLSEGLPIIATII-----DGVSLGAIAK-----ELSEINDPLLR 1057
DB 1328 LOTI---CPSITIIKKTEYVGHQLTNMTAVSVVVPYDFNSLGAINSIDKSDVDPALH 1384
QY 1058 QBIE-AKIGIMAVNUTAASTAVTSAL-----GASGFSILL 1093
DB 1385 TIVEQAKYNLLSWNDFYNTHASIMDTIARQHKSTNIEPHQSLDFDRDSKGLGLSLY 1444
QY 1094 VPLAGISAGISLVNN---ELLQDKATKVIDYFKHISLAETEGAFLLDDKIMPODD 1149
DB 1445 LDTGGYGGYQKLRNIDTASTLYOTK-----YNDNLKSNRODFFURKTQRIITWNE 1498
QY 1150 LVLSBIDFNNSITLGKCEIWRAGSGHTLDDI--DHPFSSPSITYRKPLWLSIYDVLN 1207
DB 1499 L-----GNRLKNAQLEVLK--DPIITGLIYQRRISSLLITTEYHSLAQQLSS 1548
QY 1208 IKK-----EKIDFSKOLMVLNAPRVFEGEMGTGPRSLDNDGTKLLD 1252
DB 1549 FWRVTDNPFHGHCDPHSLAQLTFIKNI-----TSNR-----NFSLSYSGSI----- 1589
QY 1253 RRDHY-EGQFWRVAFIADALITKLPREDIVNRINLG--NTRSFIVPVITTEQI 1308
DB 1590 -KVIFSESLSNNWKYIKLPLVQTSLLRDYLTTPKLSSTGGSLNIMHLPV----- 1642
QY 1309 RKNLSVSF-YGSGGSVLSLSLSPYNNIDNLNVENTWIDVNVKNTITIESDEIQ----- 1363
DB 1643 -----SFIYDIGV-----INGNRISST---DVNKIRSRUKINGDILQHOVN 1682
QY 1364 -----KGELIENILSKINIEDNKI-----ILN-----NHTINFY- 1392
DB 1683 THYLSEEQTKIKDIVDFIGIODNTIKVLESDIKPISEIQOPLHSILSROKEHVKNLLS 1742
QY 1393 GDINE-SNRP---ISLTFSEILEDINI-----IIBI---DLVSKSYKILLSGNCMKL 1436
DB 1743 GULDEFSNKLKQGLSLKTNVLSVNNFKESKINSPTVEVTVDLQGLRYVDIDTRVIGL 1802
QY 1437 I--ENSSDIQOKIDHGFNGEHQKVPYSYIDNETKNGFIDYSK---KEGLFTAEPN 1490
DB 1803 TPKEGINSISEALEHMN-----IDAIMSVIGLQVYARWIKONNINISAIIDAG 1849
QY 1491 E-SIIRNIWYPSNNLFTYSSKDLIRINKGVYK-LLIGNYFKDDMKVLSFTIEDTN 1548
DB 1850 AVSDIKNIV---DKFLGGILTLTNRYNPGVSGASLEGFISGSEVCAS----- 1897
QY 1549 TIKLANGV---YLDENGVAQILKFMNNAKSALNTSNLMFNLESINIKNIFYNLD----- 1600
DB 1898 --RMGCTAGRYL--SNVAKVIL-----PLLDI--GINWSLYDSSLSNEAKAT 1939
QY 1601 PNTEFI---LDNPF-IISGNSIGOFPE---LICDKKNIQYFFINFKETSYLYV-- 1650
DB 1940 TQIEYISTAIDVSFSSINTALSIGAAYPPLAIA-----IVBITI-FSEVKNYAVYVNO 1993
QY 1651 -GNRONLIVEPSVHDD-----SGNISSTVINFSQKLYG---IDRYNKKVII 1694
DB 1994 INERHKLWEAEKYLDNGSAKVLINKATGIIDLNNQVLGNLYDMRENPPIL 2047

```

RESULT 12
O82916 PRELIMINARY; PRT; 3169 AA.
AC O82916;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Toxin B.
GN TOXB.
OS Escherichia coli O157:H7.
OC Plasmid pO157.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7;
RX MEDLINE=98290540; PubMed=9628576;
RA Makino K., Ishii K., Yasunaga T., Hattori M., Yokoyama K.,
RA Yutsudo H.C., Kubota Y., Yamaichi Y., Iida T., Yamamoto K., Honda T.,
RA Han C.G., Ohtsubo E., Kasamatsu M., Hayashi T., Kuhara S.,
RA Shinagawa H.;
RT "Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an
RT enterohemorrhagic Escherichia coli O157:H7 derived from Sakai
RT outbreak.";
RL DNA Res. 5;1-9(1998).
DR EMBL; AB011549; BAA31815.1; .
DR PR; T00296; T00296.
DR GO; GO:0046821; C:extrachromosomal DNA; ISA.
DR InterPro; IPR007577; Gly transf. sug.
DR InterPro; IPR006473; Peptidase C58_yt.
DR Pfam; PF04488; Gly transf. sug; 1.
DR TIGRfam; TIGR01586; yopT_cys_prot; 1.
KW Plasmid.
SQ SEQUENCE 3169 AA; 361943 MW; D9BED726A62DDBFF CRC64;

Query Match 5.9%; Score 507.5; DB 2; Length 3169;
Best Local Similarity 20.08; Pred. No. 7.1e-11;
Matches 420; Conservative 338; Mismatches 716; Indels 621; Gaps 105;

QY 14 YVKPRI---QDEYVAILNAL--BEVHNMSESSVVEKYKLKLDINNLNLYNTYKSGR 68
DB 160 YIKRTRGAEDTITITQSLIINELLNGVDRNTI-PFQKISELNDIHSYENMQIKNGR 218
QY 69 -----NKALKKFE-----YLTVEVLEKNSLTPVEKNLHPFWIGG 105
DB 219 KGIEILYKQGLSSLLINVKNGKQLSDNASKIINLLGIEYQSHKV-DIEPFIHAVVAG 277
QY 106 QINDTAINYINQMKDVNSDYTVKVFYDSNAF-----LINTLAKT----- 144
DB 278 APPDNTFSYITAFUNTYKYDTYLLWIDPNAGAAKPSGILKNITAMNYAIMRLRRTNPHLA 337
QY 145 -----IVESATNNTLE--SPRENLDNPEFDY-----NKFYRKREIITDYKQHF 186
DB 338 EEMNEVLKIQNTQNEIIEKEFERLEKLENYKSLTSETKEFNVFFLESIMGNQDNY 397
QY 187 IDYYSQIENPEFI-----IDNIK--TYLSNEYSKDLKALNKYIEESLUNKITANNGN- 238
DB 398 FTYCISNGISNTDDISRLDPLTNVCLKSPVQNDFKSVKKNKRDILLKNTISQKFGDR 457
QY 239 -----DIRNLEKFADEDLVRNLQNLVERNLAAASDILRISMLKEDGGVYLDVLLPGIO 294
DB 458 FQLRDLINTLESFKKQPQDYFFQYQEMLLRWNYAAASDQVRNILKEYGGIYTDIDLAYS 517
QY 295 PDLFSKINKPDSITNTSWEIMKLEAMKYK--EYIPG--YTSKNFMDLDEVOQRSFESAL 350
DB 518 DKVQCIINE-KSDDKRFEDLKURRIISELSILKKEKYSIKH-DGLDITLNLNLL 575
QY 351 SSKSKDSBIFLPLDDIKVSPLEVKIAFANNSVI-----NQALISLKDSY 394
DB 576 -SEIEK---LTIDY--FKPVETKVRDTPKIPKRYQKWNTWNIRGNNNFMLTHKGSK 629

```

```

QY 395 CSDLVINOIKRYKILNDNLNPSINEGTDFTNTMKIFSCKLASISNE-----DN 443
DB 630 CIDFILSGQKQVLEL-QRIRDNISYNNLFYTT-----EDLSLNVAIGGIPAKKYLEH 683
QY 444 MMFMKIITNLYKGVFADPVRSTINLSGP----- 471
DB 684 GLP-----SEYRQDGTIPYVYSTINISGPDIMMQMKKYKSLGRIGEVHIKDNKLSVNF 739
QY 472 -GVYTGAYQDLLMP---KONSTNIHLLPE-----LRNEFFPKTKISQLT---BOEIT 517
DB 740 LGVYASSNKNKSNFNLNPNVSGINDITPDDESSWAVRNNDINKILFEKLNCHVPEKLPT 799
QY 518 SLWSEFNQARAKSQPEEYKGVFEFGALGEDONLDFQONTVLDKDVYKSKILSSMKTRKEY 577
DB 800 SLY-----YEIDSRSFQGW---DNKIKAVTEINKLI--KDNILLTSSNID 843
QY 578 IHVIVQLQGDKISYEASCNLFSDPYSSILYQKNIEGSETAYYYYVADAEIKEDKYRIP 637
DB 844 VKLILKL--DRELYAISKI-----DNPLALRSIRTLQQLANYVTSNTFEPENTINFI 895
QY 638 YQISNKNR-----IKLTFIGHGKSEFNTTTPANLVDVS-----LSSEIE 676
DB 896 YDFYRKQDQLLSAIDL-----FSRNDADTKIIVNYSVMKENVFLREVIS 941
QY 677 TILNLAKADISPKYIEINLLGCNMFYSISABEETYPGKLLKIKDRVSELMPSISQDSI- 735
DB 942 CVLRSKKYD---SYINEN-----KKNLSKEDA--GALRDYAKLKWKELFMSLDDDOYK 989
QY 736 -TVSANOVYEVRIEENEGKREILDHSGKWINKESIIKDISKEYISFNP----- 782
DB 990 KIITNAY---IKERDKL-----SGIYNIENSIISGHESFDIIRSHQHEWGLDSTVEQF 1041
QY 783 KENKIIVKSKYLHELSTLQBIERNNANSIDILEKKVMLTECEINVAS-----NIDQIV 837
DB 1042 KKFEFYVKSE-LSSAKSIFDDIKNK-YITDPETKRNVLVHQDSDIKERIAFLDISHVAY 1099
QY 838 EGRIEEAKNLTSDSINIYKNEFKLIESI-----SDSLYDLKHQNG 877
DB 1100 PGSLEKLQLS---GVVFSIDINIIAEYLLASYGVSHYGVVYPAPSKLLELRHT 1155
QY 878 LDDSHIFISFEDISKTENGFRIRFINKETGNSI---FIETEKEIIFSEVATHISKISNIKD 934
DB 1156 KSNSEWI--EKITP---YVVDILSDNVNVLRLPSEEBQKKILNDIKLEISKVS---E 1206
QY 935 TIFDNVANGKVKYNLDAAEHVNTLSNAFFIQSLIEYNTTK-----ESLSNLSV--- 983
DB 1207 QYFMKLTQKSSVIGIKYVDFDRYENMLFSLPINQNLTLPFMYRYFEMLYDIHIGIE 1266
QY 984 --AMKVQVYQQLFSTGLNTITDASKVVELV-----STAL-DETID 1020
DB 1267 NKANREFIYSKFPSSINLDFLINDERVLNLEGLIKKYKYLSEIHRFTLNSTSPADISIP 1326
QY 1021 LLPTLSEGLPIATII-----DGVSLGAAIK-----ELSETNDPLL 1056
DB 1327 LIQTI---CPSITITIIKKTEYVGHQLTNMTVASVVKFYDFSNLGAINSIDKSVSDVAL 1383
QY 1057 RQEIE-AKIGIMAVNLTAATAIVTSAL-----GIASGPSIL 1092
DB 1384 HTIVEQAKYNLKSWMNDFYNTHASWDITARQHKSTNIEFHPQSLFLDRDSKGKCLGSL 1443
QY 1093 LVPLAGISAGIPSLVNN---ELILQDKATKVIDYFKHISLAETEGAFLLDDKIIMPQD 1148
DB 1444 YLDTGGYGGYQKLEHNDITASTLYQTK-----YNDNLKLSNRDDFFLRKTQRIITMSN 1497
QY 1149 DJVLSEIDFNNSIITGLKEIWRABGGSGHTLTDI---DHFSSPSITYRKFWLSIYDVL 1206
DB 1498 EL-----GNRLKNAQLEVLK---DPILTGLIYQRRISILLITTEYHSLAQOIS 1547
QY 1207 NTKK-----EKIDPSKDLMTLPNAENRVFGYEMGTGPFGRSLDNDGTLL 1251
DB 1548 SPWRVTDNPFNGHCDPHSLAQALTFTKNT-----TSNR-----NFSSLYGSGI--- 1589
QY 1252 DRIRDHY-EGQFYWRYFAFIADALITKLKPRYEDTNVRINLDG---NTRSFIVPVITTEQ 1307

```

```

Db 1590 ---VKIYFSESLNNWKYIKLPLVQIGSLRLDIYLTTPKLSSTSGSLNMGHLVPV----- 1642
QY 1308 IRKLSVSP-YSGSGSYSLSPNNMIDNLVENDTWIDVNVKNITBDEIQ--- 1363
Db 1643 ---SFIYDIGV-----INGNRISST---DVQNKIRSLKINGDILQHYI 1681
QY 1364 ---KGBLIENILSKLNIENKI-----ILN---NHTINFY 1392
Db 1682 NTHYLSSECTQKIDIVDFLGIDQNTIKVKLESIDKPSBIIQOPLHLSLRQKHEVKNLL 1741
QY 1393 -GDINE-SNRP-----ISLTFSTLEINI-----IIEI---DLVSKSYKILSGNCKM 1435
Db 1742 SLLDEPNKLRKQGLSLKTWVSNVNFKEKINSQDTEVTVDLQGLYFVDIDTRVIG 1801
QY 1436 LI--ENSSDIQKIDHIGFNGEHQKYPISYVIDNETKNGFIDYSK-----KEGLFTAEFS 1489
Db 1802 LTFKEGINSLSALEHMN-----IDAIMSVIGLVQVARMIKMDNISADHA 1848
QY 1490 NB-SIIRNIYMPDSNNLIYTSKDKJRIINKGDVK--LLIGNYFKDKMKVLSFTIEDT 1547
Db 1849 GAVSDIKNIV-----DKFLGGILTLTNRVNYPGVSGASLEGFTSGLEVCA$----- 1897
QY 1548 NTKLNGV---YLDENGVAQILKFNNNAKSALNTSNLSMNFLESINIKNIFNNLD--- 1600
Db 1898 ---RMGGTAGRYL--SNVAKVTL-----PLLDI--GINWLSYDSSLNHAKA 1938
QY 1601 -FNIEFI---LDTNF-IISGNSIQGFE-----LICDKDKNIQYPIFNKIKETSYTLV- 1650
Db 1939 TTQIEYISTAIDVSFSSINTALSIGAIYPLAIA-----IVPITI-PSHEVKNVAVVYN 1992
QY 1651 ---GNQNLIVPSYHLD-----SGNISSTVINFSQKLYG-----IDRYNVKVII 1694
Db 1993 QINERHKLWLEAKYLDNGSAKVL$INKATGIIDLSNNQVLGNTIYLDRENPPIL 2047

RESULT 13
Q8EWP8
ID Q8EWP8 PRELIMINARY; PRT; 3317 AA.
AC Q8EWP8;
DT 01-WAR-2003 (TREMELrel. 23, Created)
DT 01-WAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Predicted cytoskeletal protein.
GN MYPEI550.
OS Mycoplasma penetrans.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=28227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HF-2;
RX MEDLINE=22354719; PubMed=12466555;
RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
RA Yoshino C., Horino A., Saiba T., Sasaki T., Hattori M.;
RT "The complete genomic sequence of Mycoplasma penetrans, an
RT intracellular bacterial pathogen in humans.";
RL Nucleic Acids Res. 30:5293-5300(2002).
DR ENBL; APC04170; BAC43946.1; -.
KW Complete proteome.
SQ SEQUENCE 3317 AA; 385840 MW; 0641151BAFA992BB CRC64;

Query Match 5.7%; Score 494; DB 16; Length 3317;
Best Local Similarity 20.1%; Pred. No. 2.5e-10;
Matches 419; Conservative 355; Mismatches 701; Indels 610; Gaps 105;

QY 14 YKFRIC--EDEVYALNALEYHN-MSESSVVEKYL-----KLKD-----INNLTD 57
Db 583 YINVRDLKLERQFSKIANMLEPIANKLSGDSVNEENFGQFEGRKINKIDDLISKIASERE 642
QY 58 NYLNTY-----KSGRNKALK--KFKEY 78
Db 643 NYENTYSSVILEVNAEAKENVLNLIKTEBEKNLISESKSAQNSKLGELEL 702

```

```

QY 79 LTME-----VLEKXNSLTPEKNL-----HPITWGQINDPAINYNQWKD--- 120
Db 703 IGLQNEETESLDDKXNQFISDVEQLLIEKREKLNDKILDKNSINDVMLNFKNEKINDLE 762
QY 121 -----VNSDYTVKVPYDSNAFLINTLTKT-----IVESATNWTLESFRENLDNPE 165
Db 763 YGFDRLKATINLSDLATNENQKELFNKIBETIVNNYQIVKEEVKQDLTOILNNVNLN 822
QY 166 PYNKPYKRMELIYDKQKHFDYVYSOIEENPEFIIDNIITKLSNYSYKDLKALNKYI 225
Db 823 EAYGSLNQGW--TYDKGFNDID---SQFEQ-----FRYLLKDVNTNL--INKNT 865
QY 226 E--ESLKITANNNDIRNLEKFADEDLVRLYNQELVERWNLAASDILRISMLKEDGV 283
Db 866 EYDENANSVNNISDNFKDLISGIKEENVK-YSEDIK--NVAETNLRNKGIEFGEI 922
QY 284 YLDVILPGIQPLPKFSINKPDSITNTSWEMTKLEAIWKYE---YIPGYTSKDFMDLDE 340
Db 923 ---SEKISNLVADIEKNKVAESNQANTINYILETNLANQKELVGLIKNIESNDKFLKE 979
QY 341 EVORPESALSSEKSEIFPLDDIKVSPLEVKIAPANNVINOALISLKSYSQSDVI 400
Db 980 VDKRHR-----LDNLKLNELDELISLQNSEI--QALIEKNDFIHDLEV 1022
QY 401 NOIKRKYILNDNLNPSINEGTFNTTKIPSDKLASIS-NEDNMFMKI----- 450
Db 1023 FILEKKNFNDENINELKQSAEI---TELLNKNTEINLEINFENRLREIDHFSKSELG 1079
QY 451 -----TNY--LKVGFAPDVRSTIN-----USG 470
Db 1080 NQNDPRNLEQQLTSNYEIIKNELNNDIVQVINNVNELSSVSVDRIENLSFKDKSKPDS 1139
QY 471 PGVYTGAYQDMLFKONSTNIHLPELANPEFPKTKISQLT---EOETISLWSFNQARA 527
Db 1140 IDIQDSFRHL--KDVTE--LIDKNSNYENINNVNMDLNTKPESLITKLSERNRTV 1195
QY 528 KSQFEVYK-----GYFEGALGE--DONLDPQAVNTLDKDVSKKILS 568
Db 1196 RKLNKEIKTNRNFKNIDINSIEIGDRFANLIDEIKQKNFP-ENSLDVY---SIE 1251
QY 569 SMKTRNKEYIHY-----IVQLQDKISYZEASCNL-----FSKDPYSYL 607
Db 1252 ENKSHOREILNYIKLEENNEFISQIEAKQDNE--NLKLEBELISLQNEETESL 1308
QY 608 YOKNIEGSETAYVYVADAEIKEDKIRIPYQISKNRNIKL-TFIGHGKSEFN----- 559
Db 1309 YERN-----DFINDIQAALLNKN--KLDNENIRLEDSINNA-GNFNDKISELE 1356
QY 660 ---TDTFANLDVDSLSEIETILNLAKADISPKYIEI-----NLLGCNFSYSISAEFY 711
Db 1357 QNFNNKLANIDLSLSEID-----GKQVEFFDSVQNSLSHN---YELVKNE-- 1400
QY 712 PGKLLLKIKDRVSELMPSISQDSITV-----SANQYEVRIEEGKREILD---HS 758
Db 1401 -----FKDLSLSLUNLVNVTVENTNSKIGDISSEIRDISLNNQAKFSDIDTQFDS 1453
QY 759 GKWINKE--ESII-KDISKEYIS-----FNPXENKIIYKSKYLHELSTLLQEIERNANS 810
Db 1454 FRHLKEVTENLIDKXSSNVENINIMDDINTPEFSLITKLR--NENNRVTKLNNEIKK 1511
QY 811 SVIDLEKKVMLTECEI-----NVASNDR---QVEGREIEBAKULTSDSINYI 855
Db 1512 TNLRTNKKLNTTISEIGDKFSDLIQIKQKNKNFESSQLDVIYSIIEENKSHQREILNYI 1571
QY 856 K-----NEF-----KLI-----ESISDSLYD-----LK 873
Db 1572 KNLDENNNEFINQIEAKYNNENKLEBELISLQNEETESLLEYKKNDFINEVETFLN 1631
QY 874 HONGLDSDHFI$EDISKTEGFRIFNKETGNSIFIEKEIFSEYATHI-----SKE 928
Db 1632 KKKLDNENINLEED---SVNNAIVFNK-----ISALEENFNSRLANIDLSLSEE 1681
QY 929 ISNIKDTIFDNV-----NGKLVKKNVLDAAHEV-NTLNSAFFIQSLIEYNTTKESLNL 981

```

Db 1682 IDGQVFPDVSQSSLSYVNLVKNFQKDSLNILNIN-----EVNTNVENVTK 1732
Qy 982 SVAMKQV-----YAQL-----FSTGINTITASKVVELVSTALDETIDLLPTL 1025
Db 1733 IADISTEIKDINLNQSRFSDSOFNFTLLKEVTEG--FINKENSHYDELANNITSI 1790
Qy 1026 SGLP-IATIIDGVS-LGA-AIKELSENDDLLQETEAQIGIMAVNLTAATAIVTSA 1082
Db 1791 DSFKGLISFQKSTNQIGENLVKDKHNSKHL---AIKNDSSIMEN-----1835
Qy 1083 LGIASGFSILL---VPLAGISAGIPSLVNNELIQDKATKVIDYFKHISLAETGAPT 1138
Db 1836 ---SKFDLIEBMKQTKDIENSVDTLN--WLIESKNQOEIENFNVSLLE-KSTSSL 1888
Qy 1139 LDD--KIIMPQDVLSE-----IDFNNSITLQKEMRAEGSGHTLTDIDHFPSSPS 1192
Db 1889 VNEMQAQLDKKIKLKELETLIDLQNSEI-----ESLIDDKNDPIE---1930
Qy 1193 IYRKPSWLIYDVNLTKKIDPSKDLAV-LPNA-----PNRVFGVEMGTGPRSL 1243
Db 1931 -----EIEFNQKLE--FNKENFVKLENSIGNIISAFDKKITDLEFNFDRKLSNL 1980
Qy 1244 DNDGTLLDRIRHYEGQFWRVYAFIADALITKLPYEDTNVRLNDGNTSRPIVPI 1303
Db 1981 DLSLKNLFDKDD-----STLASVINSVEDNLNSNVEILNELKELN-----2023
Qy 1304 TTEQIRKNSYSFYSGGYSLSLSPYNNIDNLVENDTWID-VDNVKNNTIESDEI 1362
Db 2024 ---KILEDI-----SEANSINEINEQPHKLEEVNSQDFQDRVLKVTW--EDII 2069
Qy 1363 QKGE-----LIENILSKLNTEDNKILNHTTIPYGDINSRPSITLFTSILEDINIIIEI 1418
Db 2070 HKSENQIVIDDLLEKINDEFRLIQNLASENELSNELMDKFNELNSGLVERFN--SNF 2127
Qy 1419 DLYSKYKILLSGNCKLIENSDDIOQ-KIDHIGF---NGEHOKYIPYSYIDN-ETKN 1472
Db 2128 DNINNKFEFIA---KLDEKTSFENRMDLNFVYELNKSYQKNIS-SYIQKIESNN 2182
Qy 1473 GFI-----DYSKKEGLTAETFSNESIIRNYMPDSNN-----LFIYSSKOLK 1514
Db 2183 EFISDLDARTLVENDKLKELEVLIDIQNEIVRLI---DENNKYIDEVEYELTNKKEAL 2239
Qy 1515 DIRIIN--KGDVKLLIGNPKDMKVSLSFTIBDTNTIKLN--GVYLDE-----1559
Db 2240 DNKKLNDIESSINNINNFEEKIKLGENFN-EKVDSIELQNLGNLNEESDKKTLDVIRS 2298
Qy 1560 -----NGVAQILKFMNNAKAL-----NTSNLSMLNTLESINIK-NIFYNN 1598
Db 2299 QLSKNVELLESEEFKNFASLIKSIENLYSSYGKDSLEQIASLKKFCYBINEGSLIGS 2358
Qy 1599 LDPNIEFI---LDTNFIISGNSIGOFELICDKD---KNIQPYFINFKIKTSYTLVYG 1651
Db 2359 INKNFDFLIKELETKQKILASQVDMNLIQSNKINSKEIVSYENLKNYANUSQKEK 2418
Qy 1652 NRQNLIVPSYHLDDSGNISSTVI-NFSQKLYGID---RYNKNV 1692
Db 2419 EKQALIFKIFDELNEINIIONQIGSLLEEDYIIDEVQKFTNKI 2463

RESULT 14

Q814R2 PRELIMINARY; PRT; 2792 AA.
ID Q814R2
AC Q814R2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Reticulocyte-binding protein, putative.
GN PFL2520W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung B., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyse S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian L.M., Mungall C.,
RA Venter J.C., Carucci B.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.,
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511(2002).
DR EMBL; AB014852; AA036586.1; --
SQ SEQUENCE 2792 AA; 328943 MW; 2952C3F104237112 CRC64;

Query Match 5.6%; Score 489; DB 5; Length 2792;
Best Local Similarity 20.1%; Pred. No. 3.1e-10;
Matches 407; Conservative 322; Mismatches 685; Indels 614; Gaps 96;

Qy 1 MLNVKAQLOQM-----VVFPRIODEYVAI-----LNAL 31
Db 856 LSTYNSTELKRLNFFSEKDELDLVESKVKQDENIFIEKNKIFDDIHKIDLYNKKTNAI 915
Qy 32 EYHN-----MSSESVVEYKVLKD--INNL-----TDNLYNTYK---KSGRNAL 72
Db 916 KNLNALINGMNNLSLDIVMKNGKDIINELSORSYLIQTNDFIDYIEKIFLKNLKNGL 975
Qy 73 KKFXYLTMEYL-----ELKNSLTPVEKNL-----HFIW--IGGOIND 109
Db 976 EEENRSLNTYMKELKEAKQEKYKLEKINITYDDTFLEKLGDNVEWVLKIELNG 1035
Qy 110 TAINY-----INQMKVNSDYTVKVFVDSNAFLINTLKKTIVESATNNTLESFR 158
Db 1036 LNVNYILQANIDTLIKPID-HIDHISLIESLKHNIENKIKVPLNERLAKDFQTK 1094
Qy 159 ENLNDPFDYKVRKMEIYYQKHPIDY-YKSOEENPEFIDNIITKILSNESYKD 217
Db 1095 FNTNDIKLDHNLITIRI---DNR---DYNHMKLEEKEDLFKNI-----NDKKEE 1140
Qy 218 LEALNKYIESLNK-----ITANNGNDI-----RNLEKFADE 249
Db 1141 IEKLLKLEENNNKMENTNISYGIKKKNLIDIESMSLSLKSITTDENLYKLQNV 1200
Qy 250 DLVR-LYNQELVERWNLAAASDILRISMLKEDGGVLDVDILP---GIQPDLPKSIKPD 305
Db 1201 DEFRLLELILIEINIKNOIKEASVERVEIYKEKILSMNSIKDISK-----1254
Qy 306 SITNTSEMIKLEAIMKYEYIPGYTSKNFDMLEEVQVSFESALSOKSEIFLDD 365
Db 1255 -LTFPKYNEIYDECTKNNIKELYT-KSSSLLEE-----SCRDKN---MDI 1296
Qy 366 IKV--SPLE--VKIAPANSVINQALISLSDYSCLVINQIK-----NRYKILN 411
Db 1297 IKTNKSVDLEYLKTSIQNNNDIKNSLASLKNMYA---ILOSIKLDVATKYILDRSYKC--1351
Qy 412 DNLAPSINEGTDFNTTKIPFSDKLASINEDNMFMKITNYLVKVPADVRSTINLGGP 471
Db 1352 -----EDHARDLEMELEKESALSKN-----IK---LKIEKAEYRNVK---1386
Qy 472 GVTYGAYQDILLMFKDNSTNIHLEPELNFNFPFKTKISQLTQEIBITLSWSP--NQARAKS 529
Db 1387 -----LGSEKHESIDY-----IFNIEKIQDISVIESDITMCIONAYDNKTSML 1432
Qy 530 QFEYKGY-PEGALGEDNDLDFQNTVLDKDYVSKKILSSMKTRNKYIHYIVLOQDK 588
Db 1433 HFQNVHRGIDLIGILNKKN-----QGVISKPAESGNI-----SEBY-----EQGN 1472
Qy 589 IS-YEASCNLFSK---DPYSSILYQKNTEGSETAYVYVADAEIKEDK-----633
Db 1473 INDYQCKCKTYSEASDNYDEISVRKD-----SLEFEKKITNLNDVLI 1517

Db 710 NKIQNMETVTKSHLKNVETNNKLSSETILDIKIYGEITNELNKTLEDFKNKEGLSNK 769
 Qy 246 ---FADEDL-VRLYNQELVE---RWNLAAASDILRLSMKEDGGVYLD-VDLPGIQPDL 297
 Db 770 IDEYAKENQVNLVYKSNILEIKKHYNQINIDINKEAKQYDQFKEHMKTIPIENENEI 829
 Qy 298 FKSINKPDSITNTSWEMIKLEAMKYKEVIGPYTSKNPDMLEEVQRPESALSCKSKS 357
 Db 830 SKTINE-----IKIMKDEFUSKVKY-----NDFKVKYKVESEHNFETELTNK- 874
 Qy 358 EIFPLDDIKVSPLEVKIAFANNVINQALISLKDSY-----C--SDLVI 400
 Db 875 -IKTEVSDSEIKKYENKFN-DSKSLINEIKTSIEBEYQINILKKVDDVIKVCNLTNELI 932
 Qy 401 NQIKRYKILNDNLNPSINEGTDFTMTKIFSKLASI----- 438
 Db 933 TNCHNKQTTLDKLNQNIKITETNSIDKIYTDKFNILTDKXTELETKFTGLSLNNHES 992
 Qy 439 -----SHEDNMFM-----IKITVYLVGFPADVRSINILSG 470
 Db 993 NKKELTYFYDLKANLGNKNENLYKQFNKEKAVEDEIKKQV-----DINKIVSNE 1045
 Qy 471 PGVVTGAYQDLMPKONSTHILEPELRNFBPPTKISQLTBOETSLWSFNQARAKSQ 530
 Db 1046 ITIYTSIYN---INEDTENEIG-----KSIELLNTKVLEKVKANTNL-----NE 1087
 Qy 531 FEYKKGYPFGALGEDNDLFAQNTVL---DKDYVSKILSSMKT-----RNKEYIHYIVQ 583
 Db 1088 IKEKLDYDFQDFGKEKNIKYPDENKIKNDIDITLQKIDKSIETLTKNGSE--NHIDE 1145
 Qy 584 LOG--DKISYEASCNLPSKDPYSSILYQKNIEGSETAYYYVADABIKIDK--YRIPYQ 639
 Db 1146 IKGQIDKXKVPKNTMENEDE-----KEIEKKIENIVEK 1179
 Qy 640 ISNKNKILFIHGKSEFNFTDIFANLVDLSASEIETILNLAKADISPKYIEINI---- 695
 Db 1180 IDKKNIT-----YKIEDKLLNEISKIEN-DKTSLE-KLKNINLSYK 1219
 Qy 696 -LGCNMF-----SYSIABETYPGKL-LLKIKDRVSLMPSISOD----- 733
 Db 1220 SLG-NLFLQIDBEKKAETHIKAMEAYIDDLNKKKSOEIEKEMWINMDIKWDIKHEM 1278
 Qy 734 -SITVSANQYEV-----RINBEGREILDSHGWIN--KEES-IKOISSK----- 775
 Db 1279 KALNISHDDYKIYHTTSKNHEEKISDIRKNSLKIIODFSEESVINDIKKEKXNVLESQ 1338
 Qy 776 -----EYIS-----FN-PRENKIIKSVKYLHELSTLLQEIRNN-----ANSSDI-- 813
 Db 1339 NNTDINOYLSKIENIYINILKINKI---KKIIDKYKEYTDEIKNKKKINAELSSEKIKIT 1395
 Qy 814 DLEKKWMLTECEINVASNIDRQIVEGREIEEAKNLTSDSIN-----YIK----- 856
 Db 1396 QLKENSLSKQCKIKTIDNYYSECIKNITNLKTYVNEKNKNINTYFKNABEYQNVNS 1455
 Qy 857 ---NEFKLIESDSLDLKHQGLD--DSHFISFEDISKTENGFRIRFINKETG-NSIF 910
 Db 1456 LNFANNIEWADTKSYIILNKKNGTNTDYNIKELKEHKKSNVYK-----DEAGKNTQE 1510
 Qy 911 IETEKEIFSEYATHI-----SKEISNIDKTIIDFNVNGKLKVKVNLDAAEHVT----- 958
 Db 1511 IKKNKELFEKYEQEVTVLLNKYAVELKNKFDKT-KNYSQIIEKIK--DAHNTFTSQAD 1567
 Qy 959 -----LNSAPFIQSLIENYTKESLSNLSVAMKVQVYAQLFSTGLNITD----- 1003
 Db 1568 KSEKKNBEIKNEQRIEDEVAKNNSKAI-LDIQLSVPEPFKIKFLKIKDLRTKSDPCLX 1626
 Qy 1004 -----ASKVVELVSTALDEITDLPTESEGLPIIATIIDGVSLGAALKELSETNDP 1054
 Db 1627 ETKOIETKISLSDITQSTKLIENKNILNTEKLESKLN--QKKNEDOKKELDEVNSK 1684
 Qy 1055 LLRQIEAKIGIMAVNTAASTAVITSALGIASGFSILLVPLAGISAGISPLVNNELIQL 1114
 Db 1685 I--KNIESNVNQHKQY-----EIGIVEKI-----NEIA-- 1711

Qy 1115 DKATKVIDYFKHSLAETGAFLLDDKIIMPODDVLSEIDFNNSITLKGKCEIWAERG 1174
 Db 1712 -KANK-----DQIESTOKLIPTIKNLISPKANDLEGIDTNKN---LGK---YNTM 1757
 Qy 1175 GSGH---TUTDDIDHFFSPS---ITYRKPWLSIVDVNLKKEKIDFSDKDLVLPNAPN 1227
 Db 1758 NNIVEEFIKSVDLITHLETVSKEPITYEQ-----IKNKRTIAQNELLTNKNVN 1807
 Qy 1228 RVFGYEMGWTPGFRSLDNDGTLLDRDRHYEG-----QFYWRY-----FAFTADAL 1274
 Db 1808 KAKSY-----LDDIEANEFDRIVTHFKNKLNDVNDKFTNEYSKVNKGFDNLSNS- 1856
 Qy 1275 ITKLKPRYEDTVNRINDGNTRGFRIVPVITTE-----QIRKNLSYSFYG 1318
 Db 1857 INNVK-KSTDENLNLINQTKEMYANIVSKYKYEAEINIFINIPKLANSLNTOIKS 1915
 Qy 1319 SGG-----SYSLSLSPY---NMNIDLNLV-----ENDTWIVDVNVVKNITIESDEIQ 1363
 Db 1916 SSGIDLKFNINIALPYLDSQKQDTLTFIPSEKTSYTKISDSYNTLLDILKRSQELQ 1975
 Qy 1364 KGE-LIENILSKLNIENDKILNNHTINFYGDJ-NESNRFISLTFSLIEDINIIEIDLIV 1421
 Db 1976 KKEQOALNLIPENELLLHDKVQATNELKDTLSLKNKKEQILNKVKLLHKSNEINLSCN 2035
 Qy 1422 SKSY-KILLSGNCMKLIENSDDIOOKIDHIGFNGEHQKYPYSIDNETKNGFI-DYSK 1479
 Db 2036 SQNYDTILESSEKYDKIKEKSNYEKEKENLGIN-----FDVKAMEEQFNNDIKDIEK 2087
 Qy 1480 KEGLFTAEFSNESIIRNIYMPDSNNLIYSSKOLKDIRIINKGVDKLLIGNFYKDDMKVS 1539
 Db 2088 LENNY-----KHSEKDNVNFSEENNTLQSKKCLKELT-----NAFNAEIK-- 2128
 Qy 1540 LSFTIEDTNTIKLGVYLDENGVAQILKFMNNAKSA-LNTSNSLMNPLESINIKNIPYNN 1598
 Db 2129 -----KIEDKIEKNGL--INKLIETRXDCMLFTYKTL---VETLKIKTTDY-- 2170
 Qy 1599 LDPNIEFTLDNFTIISGNSIGQPELICDDKXNIQPYFINFKIKETSYTLVGNRQ--NL 1656
 Db 2171 -----TKFITSATKFSKEFLKYIDATSNLSNDDIN--TLQTKYDLNQINKHVASM 2218
 Qy 1657 IVEPSYHLDDSGNI-----SSTVINPSOKLYGIDRYNVKV 1692
 Db 2219 VADAT---NDNNNLIEKEKEATKTINNLT--LFTIDS--NKI 2254

Search completed: April 1, 2004, 16:44:42
 Job time : 92.75 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 16:34:35 ; Search time 17.5 Seconds
(without alignments)
5058.244 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1700

Perfect score: 8675

Sequence: 1 MNLVNKAQLQKVVYKFRIQ.....YLYGIDRVYKVIAPNLYT 1700

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6759.5	77.9	2366	1	TOXB_CLODI
2	4467	51.5	2710	1	TOXA_CLODI
3	395	4.6	2867	1	RBP2_PLAVB
4	360	4.1	2869	1	RBP1_PLAVB
5	312	3.6	1639	1	MSPI_PLAFW
6	311.5	3.6	1726	1	MSPI_PLAF3
7	309.5	3.6	1682	1	MSPI_PLAF3
8	309.5	3.6	1957	1	SPOF_SCHPO
9	309	3.6	1630	1	MSPI_PLAFK
10	307.5	3.5	1701	1	MSPI_PLAFF
11	306.5	3.5	1726	1	MSPI_PLAFF
12	304.5	3.5	1701	1	MSPI_PLAFM
13	302.5	3.5	2136	1	YCF2_MARPO
14	289	3.3	3433	1	UTRO_HUMAN
15	288	3.3	1875	1	MLP1_YEAST
16	288	3.3	2748	1	NUM1_YEAST
17	285.5	3.3	2894	1	YD96_METJA
18	283.5	3.3	1956	1	ATX1_PLAPA
19	278	3.2	2059	1	TEGU_HSV7J
20	278	3.2	5171	1	BPEA_HUMAN
21	274	3.2	2339	1	RPCI_PLAPA
22	273.5	3.2	1928	1	MYSL_YEAST
23	272.5	3.1	6885	1	SNE2_HUMAN
24	270.5	3.1	1892	1	Y835_RICCN
25	269.5	3.1	2167	1	BEM2_YEAST
26	268	3.1	2663	1	CENE_HUMAN
27	267	3.1	1803	1	YJL3_YEAST
28	265.5	3.1	3210	1	CENF_HUMAN
29	265	3.1	4196	1	DYHC_SCHPO
30	263.5	3.0	1727	1	ALM1_SCHPO
31	261.5	3.0	2875	1	RRPL_TSWV1
32	260.5	3.0	1628	1	NAGH_CLOPE
33	257.5	3.0	1162	1	EXEN_CLOBO

34 257.5 3.0 1679 1 YIO9_YEAST P40457 saccharomyc
35 257 3.0 1790 1 USO1_YEAST P25386 saccharomyc
36 257 3.0 2014 1 XJ07_YEAST P39526 saccharomyc
37 256 3.0 3144 1 VP13_YEAST Q07878 saccharomyc
38 254.5 2.9 3135 1 S230_PLAFO Q08372 plasmodium
39 254.5 2.9 4092 1 DYHC_YEAST P36022 saccharomyc
40 252.5 2.9 8545 1 ANCL_CABEL Q9n4m4 caenorhabdi
41 250.5 2.9 1295 1 EXAL_CLOBO P10845 clostridium
42 250.5 2.9 1868 1 YH00_YEAST P38737 saccharomyc
43 250 2.9 3343 1 YOG7_CABEL P34616 caenorhabdi
44 250 2.9 4563 1 APB_HUMAN P04114 homo sapien
45 248 2.9 1274 1 BXF_CLOBO P30996 clostridium

ALIGNMENTS

RESULT 1

TOXB_CLODI
ID TOXB_CLODI STANDARD; PRT; 2366 AA.
AC P18177;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Toxin B
GN TOXB OR TCDB
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]_TaxID=1496;
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RX MEDLINE=90326540; PubMed=2374729;
RA Barroso L.A., Wang S.Z., Phelps C.J., Johnson J.L., Wilkins T.D.;
RT "Nucleotide sequence of Clostridium difficile toxin B gene."
RL Nucleic Acids Res. 18:4004-4004(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI 10463;
RA von Eichel-Streiber C.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1271-2366 FROM N.A.
RC STRAIN=VPI 10463;
RX MEDLINE=92293124; PubMed=1603068;
RA Eichel-Streiber C., Laufenberg-Feldmann R., Sartingen S., Schulze J.,
RT Sauerborn M.;
RL "Comparative sequence analysis of the Clostridium difficile toxins A
and B."
RL Mol. Gen. Genet. 233:260-268(1992).
CC -!- FUNCTION: Cytotoxin.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X531138; CAA37298.1; -;
DR EMBL; X92982; CAA63562.1; -;
DR EMBL; X60984; CAA43299.1; -;
DR PIR; A27636; A27636.
DR PIR; S10317; S10317.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR007577; Gly_transf_sug.
DR Pfam; PF01473; CW_binding_1; 17;
DR Pfam; PF04488; Gly_transf_sug; 1.
KW Toxin.
SQ SEQUENCE 2366 AA; 269709 MW; E1024BD8B8A56ADF CRC64;

Query Match	77.9%	Score 6759.5	DB 1	Length 2366
Best Local Similarity	76.6%	Pred. No. 1.6e-250		
Matches 1302	Conservative 191	Mismatches 206	Indels 1	Gaps 1
1	MSLVNRKAQGLQKVVYKFRIOEBYVAILNALAEYHNMSSESVBEKYLKLDINNLTDNYL	60		
1	MSLVNRKQLEKMANVFRIOEBYVAILDALAEYHNMSNTVBEKYLKLDINSLTDIYI	60		
61	NTYKSGRNKALKKFKXYLTMEVLELKNNSLTPVEKNLHFIWIGQINDTAIYINQWKD	120		
61	DTYKSGRNKALKKFKXYLTVEVLELKNNSLTPVEKNLHFIWIGQINDTAIYINQWKD	120		
121	VNSDYTVKVPYDSNAFLINTLKKTIVESANTNTLESFRENLANDPEFYNYFKYKRWELIY	180		
121	VNSDYNVVPYDSNAFLINTLKKTIVESAINDTLESFRENLANDPRFYNYKFFKRWEIY	180		
181	DKQKHFDYKSGIOENPRFIDNIKTIVLSNEYSKOLEALNKYIEESLNKIKTANNNDI	240		
181	DKQKNFINNYKAQRENPELIDDIKVTVLSNEYSKIDEIUNTYIEESLNKIKTQNSGNDV	240		
241	RNLEKFADEDLVRLYNQELVERWNLAASDILRLISMLKEDGGVYLDVLDLPGIQPOLFKS	300		
241	RNFEFKNGSEFNLYEQELVERWNLAASDILRLISALKEIGWYLDVDMPLGQPOLFES	300		
301	INKPDSITNTSWEMIKLEALMKYKVIPOYTSKCNFMDLDBEVQSFSAIUSKSDKSEIF	360		
301	IEKPSSVTVDFWEMTKLEALMKYKVIPIPYTSEHFMDLDBEVQSFSAIUSKSDKSEIF	360		
361	LPDDIKVSPLEVKIAPANNVINOALISLDSYSCDVLINOIKNRKYKILNDNLNPSINE	420		
361	SSLGMEASPLEVKIAPNKSIIQGLISVKSOCNSLIVKQENRYKILNLSLNPAISE	420		
421	GTDENTNTMKI FSDKLASINEDNMFMKIIITNLYKVGFPDVRSTINLSGPGVVTGAYQD	480		
421	DNDENTNTN FDSIMAEANADNGRFMEELKYLAVGFPDPVKITINLSGPEAAVAYQD	480		
481	LLMPKDNSTNHLLEPELNFPEPKTKISQLTEQISITLSWSPNOARAKSQPEYKGYFE	540		
481	LLMPKEGSMNTHLEADLRNFEISKTNISQSTEQEWASLWFSDFADARAKAQPEYKRYFE	540		
541	GALGEDDNLQFAQNTVLDKYVSKTILSMKTRNKEYTHYIVOLQGDKISYEASCNLFK	600		
541	GSLEDNLDQFNIVVDKELYLEKISSARSEKGYTHYIVOLQDKISYEACNLFAP	600		
601	DPSYSLVQKNIEGSEYAYVYVADAEIKEDIKYRIPQISMKRNKILTFIGHGKSEFNT	660		
601	TPYDSVLFQKNIEDESEAYYNPCDGEIQEIDKYKIPSIIDRPKILTFIGHGKSEFNT	660		
661	DTFANLDVDSLSSBIETILNADAKISPKYIEINLGCNMFYSYSISABETYPGKLILKIK	720		
661	DIFAGFDVDSLSTEIEAAIDLAKEDISPKSIEINLGCNMFYSINVEETYPGKLILKVK	720		
721	DRVELMPSISQDSITVSAHQYVRNREGKREILDHSGKWNKEESIKDISKEYISF	780		
721	DKISELMPISQDSIIVSAHQYVRNREGRELDSHGEWNKEESIKDISKEYISF	780		
781	NPKENKIIVKSKYLHELSTLLOERNNANSDDILEKKVMLTECEINVASNIDRQIVEGR	840		
781	NPKENKIIVKSKNPELSTLLOERNNNSDDIEELKVMTECEINVINSDIQIVEER	840		
841	IEEAKNLTSDSINYKNEFKLIESIDSLVDLKHQGLDDSHFISFEDISKTENGFIIRF	900		
841	IEEAKNLTSDSINYKDEFKLIESIDALCDLQKQNELEDSHFISFEDISDEGFSIRF	900		
901	INKETGNSIFETEKELPSEYATHISKEISNIKDTTFDYNVNGKLYKKNLDAAHEVNTLN	960		
901	INKETGESIFVETEKELPSEYANHITEISKIKGTTFDYNVNGKLYKKNLDTTHEVNTLN	960		
961	SAFFITQSILEYNTTKESLNSLVAMKVQVYAQFSTGLNTITDASKVVELVSTALDETID	1020		
961	AAFFITQSILEYNSKESLNSLVAMKVQVYAQFSTGLNTITDAAKVELVSTALDETID	1020		
1021	LLPPTLSEGLPIIATIDGVSGLGAIEKLSSETNDPLLRQIEIAKIGIMAVNTAASTAIVT	1080		

Db	1021	LFPTLSEGLPIATIIDGVSLGAAIKELSETSDPLLRQEBIAKIGIMAVNLTTATTAIT	1080
Qy	1081	SALGIASGFSILLVPLAGISAGIPSLVNNELLQDKATKVIDFYFCHI SLAETEGAFITLD	1140
Db	1081	SSLGIASGFSILLVPLAGISAGIPSLVNNELLVLRDKATKVVDFYFKHSLVETEGVFTLLD	1140
Qy	1141	DKIINPODDLVLSEIDFNNSITLKGCEIWRAGGSGHGLTDDIDHFPSPSPITVYRKPL	1200
Db	1141	DKIMPQDDLVLSEIDFNNSITLKGCEIWRAGGSGHGLTDDIDHFPSPSPITVYRKPL	1200
Qy	1201	SIYDVLNIIKKEIDFSGKOLMVLPNAPNRVFGYEMGWTGFRSLDNDGTGLLRDIRDHYEG	1260
Db	1201	SIYDVLVQKEELDLSKOLMVLPNAPNRVFAETGWTGFRSLDNDGTGLLRDIRDHYEG	1260
Qy	1261	QFYWRYFAFIADALITKUPRYEDTNVIRNLGNTRSFIVPVITTEQIRKNLSYSPFYGSG	1320
Db	1261	EFYWRYFAFIADALITLKPRIEDTNIRNLGNTRSFIVPVITTEYIRKLSYSPFYGSG	1320
Qy	1321	GSYSLSLSPYNNIDNLVENDTWIDVDNVKNITIESDEIQGELIENILSKLNIEDN	1380
Db	1321	GTVALSLSPYNNIGINIEUSESDDWIIDVDNVVRDVITIESDKIKGGLIEGLISTLSEEN	1380
Qy	1381	KIILNNHTINFYGDINESNRFSLTFSIILEDINIIIEIDLVSKEYKILLSGNCMKLIENS	1440
Db	1381	KIILNSHEINFSGEVNGSGFVSLFSLILEGINAIEVDLLSKSYKLLISGELKILMLNS	1440
Qy	1441	SDIQQKIDHIGFNGEHQKYPYSYIDNETKYNGFDIDYKKEGLFTAEPFONESIIRNIYP	1500
Db	1441	NHIQQKIDYIGFNSLQKNIPYSYFVDSSEKNGFINGSTKEGLFVSELPDVVLISKVYMD	1500
Qy	1501	DSNNLFIYSKDLKDIRIINKGDVKLLIGNYFKDDMKVLSLPTIEDTNTIKLNGVYLDEN	1560
Db	1501	DSKPSFGYYSNNLKDKVKVITKDNVNLTGYLKDIDIKISLTLQDEKTIKLSNVHLDES	1560
Qy	1561	GVAQILKEWNNAKSALNTSNSLMNPLESINIKNIFVNNLDPNIEFILDNTFIISGNSIG	1620
Db	1561	GVAELIKFMR-KGNTNTSDLSMFLESNIXISIFVNFQSNIKFILDANFIISGTTSIG	1619
Qy	1621	QFELICDCKQKIQYFINFKIKETSYTLVYGNRQNLIVPSPYHLDSDGNISSVTINFSQK	1680
Db	1620	QFEPICDENDIQYIFKNTLETNTVLYVGNRQNMIVBNPYDLDDSGDISSTVINFSQK	1679
Qy	1681	YLYGIDRYNKVLIAPNLYT 1700	
Db	1680	YLYGIDSCVKNVISPNIYT 1699	
RESULT 2			
TOXA_CLODI			
ID	TOXA_CLODI	STANDARD;	PRT; 2710 AA.
AC	P16154;		
DT	01-APR-1990 (Rel. 14, Created)		
DT	01-FEB-1996 (Rel. 33, Last sequence update)		
DT	15-MAR-2004 (Rel. 43, Last annotation update)		
DE	Toxin A.		
GN	TOXA OR TCDA.		
OS	Clostridium difficile.		
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;		
OC	Clostridium.		
OK	NCBI_Taxid=1496;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=VPI 10463;		
EX	MEDLINE=90221894; PubMed=2109310;		
RA	Sauerborn M., von Eichel-Streiber C.;		
RT	"Nucleotide sequence of Clostridium difficile toxin A.";		
RL	Nucleic Acids Res. 18:1629-1630(1990).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=VPI 10463;		
RX	MEDLINE=90129305; PubMed=2105276;		
RA	Dove C.H., Wang S.Z., Price S.B., Phelps C.J., Lysterly D.M.,		

RESIST. T. 2

TOXA CLOD

TOXA CLODI
ID - TOXA CLODI
STANDARD;
PRT; 2710 AA.

AC P16154;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Toxin A.

GN TOXA OR TCDA.

OS Clostridium difficile.
OC Bacterio. E. coli.

00 Clostridium
00 bacteria; filamentous

NCBI TaxID=1496:

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=VPI 10463;

RX MEDLINE=90221894; PubMed=2109310;

RA Sauerborn M., von Eichel-Streiber C.;

RT: "NUCLEOTIDE SEQUENCE OF CROSSLINKING AFFILIATE COXIN A:
PR: Nucleic Acids Res 19:1629-1630 (1990)

PN [2]
RL NUCLEIC ACIDS RES. 19:1025-1030 (1990)

RP SEQUENCE FROM N.A.

RC STRAIN=VPI 10463;

RX MEDLINE=90129305; PubMed=2105276;

RA Dove C.H., Wang S.Z., Price S.B.,

QY 1602 NIEPILDTNFIISGSIQFELICDKKNIQYFNFINKIKETSYTYLVGNRONLIVPS 1661
 DB 1599 NINVIDKYFTLVGKTLGVVEFCDDNNKNIDIFYGEWKTSKSTIFSNGRNVVVEPI 1658

QY 1662 YHLDSDSGNISGVNFOSKLYGIDRVYVNVKVIAPLNT 1700
 DB 1659 YN-PDTGEDISTSLDFSEYELYGIDRVYVNVKVIAPLNT 1696

RESULT 3

ID RBP2 PLAVB STANDARD; PRT: 2867 AA.

AC Q00799; Q9N2M3;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Reticulocyte binding protein 2 precursor (PvRBP-2).
 GN RBP-2 OR RBP2.
 OS Plasmodium vivax (strain Belem).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=31273;
 RN [1]
 RP SEQUENCE FROM N.A., AND REVISIONS TO 2438-2439.
 RX MEDLINE=20299192; PubMed=10838229;
 RA Galinski M.R., Xu M., Barnwell J.W.;
 RT "Plasmodium vivax reticulocyte binding protein-2 (PvRBP-2) shares
 RT structural features with PvRBP-1 and the Plasmodium yoelii 235 kDa
 RT rhoptry protein family";
 RL Mol. Biochem. Parasitol. 108:257-262(2000).
 RN [2]
 RP SEQUENCE OF 1189-2439 FROM N.A.
 RX MEDLINE=92315338; PubMed=1617731;
 RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
 RT "A reticulocyte-binding protein complex of Plasmodium vivax
 RT merozoites";
 RL Cell 69:1213-1226(1992).
 CC -!- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to
 CC human reticulocyte cells.
 CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).
 CC

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

 CC EMBL; AF184623; AAF76525.1; -.
 DR HSP; P03069; IGCM. Signal; Transmembrane; Repeat.
 KW Malaria; Receptor; Signal; Transmembrane; Repeat.
 FT SIGNAL 1 21
 FT CHAIN 22 2867
 FT DOMAIN 22 2805
 FT TRANSMEM 2806 2826
 FT DOMAIN 2827 2867
 FT DOMAIN 44 133
 FT DOMAIN 560 758
 FT DOMAIN 1112 1285
 FT DOMAIN 2758 2785
 FT REPEAT 2758 2761
 FT REPEAT 2762 2765
 FT REPEAT 2766 2769
 FT REPEAT 2770 2773
 FT REPEAT 2774 2777
 FT REPEAT 2778 2781
 FT REPEAT 2782 2785
 FT REPEAT 2786 2789
 FT REPEAT 2790 2793
 FT REPEAT 2794 2797
 FT REPEAT 2798 2801
 FT REPEAT 2802 2805
 FT REPEAT 2806 2809
 FT REPEAT 2810 2813
 FT REPEAT 2814 2817
 FT REPEAT 2818 2821
 FT REPEAT 2822 2825
 FT REPEAT 2826 2829
 FT REPEAT 2830 2833
 FT REPEAT 2834 2837
 FT REPEAT 2838 2841
 FT REPEAT 2842 2845
 FT REPEAT 2846 2849
 FT REPEAT 2850 2853
 FT REPEAT 2854 2857
 FT REPEAT 2858 2861
 FT REPEAT 2862 2865
 FT REPEAT 2866 2869
 FT REPEAT 2870 2873
 FT REPEAT 2874 2877
 FT REPEAT 2878 2881
 FT REPEAT 2882 2885
 FT REPEAT 2886 2889
 FT REPEAT 2890 2893
 FT REPEAT 2894 2897
 FT REPEAT 2898 2901
 FT REPEAT 2902 2905
 FT REPEAT 2906 2909
 FT REPEAT 2910 2913
 FT REPEAT 2914 2917
 FT REPEAT 2918 2921
 FT REPEAT 2922 2925
 FT REPEAT 2926 2929
 FT REPEAT 2930 2933
 FT REPEAT 2934 2937
 FT REPEAT 2938 2941
 FT REPEAT 2942 2945
 FT REPEAT 2946 2949
 FT REPEAT 2950 2953
 FT REPEAT 2954 2957
 FT REPEAT 2958 2961
 FT REPEAT 2962 2965
 FT REPEAT 2966 2969
 FT REPEAT 2970 2973
 FT REPEAT 2974 2977
 FT REPEAT 2978 2981
 FT REPEAT 2982 2985
 FT REPEAT 2986 2989
 FT REPEAT 2990 2993
 FT REPEAT 2994 2997
 FT REPEAT 2998 3001
 FT REPEAT 3002 3005
 FT REPEAT 3006 3009
 FT REPEAT 3010 3013
 FT REPEAT 3014 3017
 FT REPEAT 3018 3021
 FT REPEAT 3022 3025
 FT REPEAT 3026 3029
 FT REPEAT 3030 3033
 FT REPEAT 3034 3037
 FT REPEAT 3038 3041
 FT REPEAT 3042 3045
 FT REPEAT 3046 3049
 FT REPEAT 3050 3053
 FT REPEAT 3054 3057
 FT REPEAT 3058 3061
 FT REPEAT 3062 3065
 FT REPEAT 3066 3069
 FT REPEAT 3070 3073
 FT REPEAT 3074 3077
 FT REPEAT 3078 3081
 FT REPEAT 3082 3085
 FT REPEAT 3086 3089
 FT REPEAT 3090 3093
 FT REPEAT 3094 3097
 FT REPEAT 3098 3101
 FT REPEAT 3102 3105
 FT REPEAT 3106 3109
 FT REPEAT 3110 3113
 FT REPEAT 3114 3117
 FT REPEAT 3118 3121
 FT REPEAT 3122 3125
 FT REPEAT 3126 3129
 FT REPEAT 3130 3133
 FT REPEAT 3134 3137
 FT REPEAT 3138 3141
 FT REPEAT 3142 3145
 FT REPEAT 3146 3149
 FT REPEAT 3150 3153
 FT REPEAT 3154 3157
 FT REPEAT 3158 3161
 FT REPEAT 3162 3165
 FT REPEAT 3166 3169
 FT REPEAT 3170 3173
 FT REPEAT 3174 3177
 FT REPEAT 3178 3181
 FT REPEAT 3182 3185
 FT REPEAT 3186 3189
 FT REPEAT 3190 3193
 FT REPEAT 3194 3197
 FT REPEAT 3198 3201
 FT REPEAT 3202 3205
 FT REPEAT 3206 3209
 FT REPEAT 3210 3213
 FT REPEAT 3214 3217
 FT REPEAT 3218 3221
 FT REPEAT 3222 3225
 FT REPEAT 3226 3229
 FT REPEAT 3230 3233
 FT REPEAT 3234 3237
 FT REPEAT 3238 3241
 FT REPEAT 3242 3245
 FT REPEAT 3246 3249
 FT REPEAT 3250 3253
 FT REPEAT 3254 3257
 FT REPEAT 3258 3261
 FT REPEAT 3262 3265
 FT REPEAT 3266 3269
 FT REPEAT 3270 3273
 FT REPEAT 3274 3277
 FT REPEAT 3278 3281
 FT REPEAT 3282 3285
 FT REPEAT 3286 3289
 FT REPEAT 3290 3293
 FT REPEAT 3294 3297
 FT REPEAT 3298 3301
 FT REPEAT 3302 3305
 FT REPEAT 3306 3309
 FT REPEAT 3310 3313
 FT REPEAT 3314 3317
 FT REPEAT 3318 3321
 FT REPEAT 3322 3325
 FT REPEAT 3326 3329
 FT REPEAT 3330 3333
 FT REPEAT 3334 3337
 FT REPEAT 3338 3341
 FT REPEAT 3342 3345
 FT REPEAT 3346 3349
 FT REPEAT 3350 3353
 FT REPEAT 3354 3357
 FT REPEAT 3358 3361
 FT REPEAT 3362 3365
 FT REPEAT 3366 3369
 FT REPEAT 3370 3373
 FT REPEAT 3374 3377
 FT REPEAT 3378 3381
 FT REPEAT 3382 3385
 FT REPEAT 3386 3389
 FT REPEAT 3390 3393
 FT REPEAT 3394 3397
 FT REPEAT 3398 3401
 FT REPEAT 3402 3405
 FT REPEAT 3406 3409
 FT REPEAT 3410 3413
 FT REPEAT 3414 3417
 FT REPEAT 3418 3421
 FT REPEAT 3422 3425
 FT REPEAT 3426 3429
 FT REPEAT 3430 3433
 FT REPEAT 3434 3437
 FT REPEAT 3438 3441
 FT REPEAT 3442 3445
 FT REPEAT 3446 3449
 FT REPEAT 3450 3453
 FT REPEAT 3454 3457
 FT REPEAT 3458 3461
 FT REPEAT 3462 3465
 FT REPEAT 3466 3469
 FT REPEAT 3470 3473
 FT REPEAT 3474 3477
 FT REPEAT 3478 3481
 FT REPEAT 3482 3485
 FT REPEAT 3486 3489
 FT REPEAT 3490 3493
 FT REPEAT 3494 3497
 FT REPEAT 3498 3501
 FT REPEAT 3502 3505
 FT REPEAT 3506 3509
 FT REPEAT 3510 3513
 FT REPEAT 3514 3517
 FT REPEAT 3518 3521
 FT REPEAT 3522 3525
 FT REPEAT 3526 3529
 FT REPEAT 3530 3533
 FT REPEAT 3534 3537
 FT REPEAT 3538 3541
 FT REPEAT 3542 3545
 FT REPEAT 3546 3549
 FT REPEAT 3550 3553
 FT REPEAT 3554 3557
 FT REPEAT 3558 3561
 FT REPEAT 3562 3565
 FT REPEAT 3566 3569
 FT REPEAT 3570 3573
 FT REPEAT 3574 3577
 FT REPEAT 3578 3581
 FT REPEAT 3582 3585
 FT REPEAT 3586 3589
 FT REPEAT 3590 3593
 FT REPEAT 3594 3597
 FT REPEAT 3598 3601
 FT REPEAT 3602 3605
 FT REPEAT 3606 3609
 FT REPEAT 3610 3613
 FT REPEAT 3614 3617
 FT REPEAT 3618 3621
 FT REPEAT 3622 3625
 FT REPEAT 3626 3629
 FT REPEAT 3630 3633
 FT REPEAT 3634 3637
 FT REPEAT 3638 3641
 FT REPEAT 3642 3645
 FT REPEAT 3646 3649
 FT REPEAT 3650 3653
 FT REPEAT 3654 3657
 FT REPEAT 3658 3661
 FT REPEAT 3662 3665
 FT REPEAT 3666 3669
 FT REPEAT 3670 3673
 FT REPEAT 3674 3677
 FT REPEAT 3678 3681
 FT REPEAT 3682 3685
 FT REPEAT 3686 3689
 FT REPEAT 3690 3693
 FT REPEAT 3694 3697
 FT REPEAT 3698 3701
 FT REPEAT 3702 3705
 FT REPEAT 3706 3709
 FT REPEAT 3710 3713
 FT REPEAT 3714 3717
 FT REPEAT 3718 3721
 FT REPEAT 3722 3725
 FT REPEAT 3726 3729
 FT REPEAT 3730 3733
 FT REPEAT 3734 3737
 FT REPEAT 3738 3741
 FT REPEAT 3742 3745
 FT REPEAT 3746 3749
 FT REPEAT 3750 3753
 FT REPEAT 3754 3757
 FT REPEAT 3758 3761
 FT REPEAT 3762 3765
 FT REPEAT 3766 3769
 FT REPEAT 3770 3773
 FT REPEAT 3774 3777
 FT REPEAT 3778 3781
 FT REPEAT 3782 3785
 FT REPEAT 3786 3789
 FT REPEAT 3790 3793
 FT REPEAT 3794 3797
 FT REPEAT 3798 3801
 FT REPEAT 3802 3805
 FT REPEAT 3806 3809
 FT REPEAT 3810 3813
 FT REPEAT 3814 3817
 FT REPEAT 3818 3821
 FT REPEAT 3822 3825
 FT REPEAT 3826 3829
 FT REPEAT 3830 3833
 FT REPEAT 3834 3837
 FT REPEAT 3838 3841
 FT REPEAT 3842 3845
 FT REPEAT 3846 3849
 FT REPEAT 3850 3853
 FT REPEAT 3854 3857
 FT REPEAT 3858 3861
 FT REPEAT 3862 3865
 FT REPEAT 3866 3869
 FT REPEAT 3870 3873
 FT REPEAT 3874 3877
 FT REPEAT 3878 3881
 FT REPEAT 3882 3885
 FT REPEAT 3886 3889
 FT REPEAT 3890 3893
 FT REPEAT 3894 3897
 FT REPEAT 3898 3901
 FT REPEAT 3902 3905
 FT REPEAT 3906 3909
 FT REPEAT 3910 3913
 FT REPEAT 3914 3917
 FT REPEAT 3918 3921
 FT REPEAT 3922 3925
 FT REPEAT 3926 3929
 FT REPEAT 3930 3933
 FT REPEAT 3934 3937
 FT REPEAT 3938 3941
 FT REPEAT 3942 3945
 FT REPEAT 3946 3949
 FT REPEAT 3950 3953
 FT REPEAT 3954 3957
 FT REPEAT 3958 3961
 FT REPEAT 3962 3965
 FT REPEAT 3966 3969
 FT REPEAT 3970 3973
 FT REPEAT 3974 3977
 FT REPEAT 3978 3981
 FT REPEAT 3982 3985
 FT REPEAT 3986 3989
 FT REPEAT 3990 3993
 FT REPEAT 3994 3997
 FT REPEAT 3998 4001
 FT REPEAT 4002 4005
 FT REPEAT 4006 4009
 FT REPEAT 4010 4013
 FT REPEAT 4014 4017
 FT REPEAT 4018 4021
 FT REPEAT 4022 4025
 FT REPEAT 4026 4029
 FT REPEAT 4030 4033
 FT REPEAT 4034 4037
 FT REPEAT 4038 4041
 FT REPEAT 4042 4045
 FT REPEAT 4046 4049
 FT REPEAT 4050 4053
 FT REPEAT 4054 4057
 FT REPEAT 4058 4061
 FT REPEAT 4062 4065
 FT REPEAT 4066 4069
 FT REPEAT 4070 4073
 FT REPEAT 4074 4077
 FT REPEAT 4078 4081
 FT REPEAT 4082 4085
 FT REPEAT 4086 4089
 FT REPEAT 4090 4093
 FT REPEAT 4094 4097
 FT REPEAT 4098 4101
 FT REPEAT 4102 4105
 FT REPEAT 4106 4109
 FT REPEAT 4110 4113
 FT REPEAT 4114 4117
 FT REPEAT 4118 4121
 FT REPEAT 4122 4125
 FT REPEAT 4126 4129
 FT REPEAT 4130 4133
 FT REPEAT 4134 4137
 FT REPEAT 4138 4141
 FT REPEAT 4142 4145
 FT REPEAT 4146 4149
 FT REPEAT 4150 4153
 FT REPEAT 4154 4157
 FT REPEAT 4158 4161
 FT REPEAT 4162 4165
 FT REPEAT 4166 4169
 FT REPEAT 4170 4173
 FT REPEAT 4174 4177
 FT REPEAT 4178 4181
 FT REPEAT 4182 4185
 FT REPEAT 4186 4189
 FT REPEAT 4190 4193
 FT REPEAT 4194 4197
 FT REPEAT 4198 4201
 FT REPEAT 4202 4205
 FT REPEAT 4206 4209
 FT REPEAT 4210 4213
 FT REPEAT 4214 4217
 FT REPEAT 4218 4221
 FT REPEAT 4222 4225
 FT REPEAT 4226 4229
 FT REPEAT 4230 4233
 FT REPEAT 4234 4237
 FT REPEAT 4238 4241
 FT REPEAT 4242 4245
 FT REPEAT 4246 4249
 FT REPEAT 4250 4253
 FT REPEAT 4254 4257
 FT REPEAT 4258 4261
 FT REPEAT 4262 4265
 FT REPEAT 4266 4269
 FT REPEAT 4270 4273
 FT REPEAT 4274 4277
 FT REPEAT 4278 4281
 FT REPEAT 4282 4285
 FT REPEAT 4286 4289
 FT REPEAT 4290 4293
 FT REPEAT 4294 4297
 FT REPEAT 4298 4301
 FT REPEAT 4302 4305
 FT REPEAT 4306 4309
 FT REPEAT 4310 4313
 FT REPEAT 4314 4317
 FT REPEAT 4318 4321
 FT REPEAT 4322 4325
 FT REPEAT 4326 4329
 FT REPEAT 4330 4333
 FT REPEAT 4334 4337
 FT REPEAT 4338 4341
 FT REPEAT 4342 4345
 FT REPEAT 4346 4349
 FT REPEAT 4350 4353
 FT REPEAT 4354 4357
 FT REPEAT 4358 4361
 FT REPEAT 4362 4365
 FT REPEAT 4366 4369
 FT REPEAT 4370 4373
 FT REPEAT 4374 4377
 FT REPEAT 4378 4381
 FT REPEAT 4382 4385
 FT REPEAT 4386 4389
 FT REPEAT 4390 4393
 FT REPEAT 4394 4397
 FT REPEAT 4398 4401
 FT REPEAT 4402 4405
 FT REPEAT 4406 4409
 FT REPEAT 4410 4413
 FT REPEAT 4414 4417
 FT REPEAT 4418 4421
 FT REPEAT 4422 4425
 FT REPEAT 4426 4429
 FT REPEAT 4430 4433
 FT REPEAT 4434 4437
 FT REPEAT 4438 4441
 FT REPEAT 4442 4445
 FT REPEAT 4446 4449
 FT REPEAT 4450 4453
 FT REPEAT 4454 4457
 FT REPEAT 4458 4461
 FT REPEAT 4462 4465
 FT REPEAT 4466 4469
 FT REPEAT 4470 4473
 FT REPEAT 4474 4477
 FT REPEAT 4478 4481
 FT REPEAT 4482 4485
 FT REPEAT 4486 4489
 FT REPEAT 4490 4493
 FT REPEAT 4494 4497
 FT REPEAT 4498 4501
 FT REPEAT 4502 4505
 FT REPEAT 4506 4509
 FT REPEAT 4510 4513
 FT REPEAT 4514 4517
 FT REPEAT 4518 4521
 FT REPEAT 4522 4525
 FT REPEAT 4526 4529
 FT REPEAT 4530 4533
 FT REPEAT 4534 4537
 FT REPEAT 4538 4541
 FT REPEAT 4542 4545
 FT REPEAT 4546 4549
 FT REPEAT 4550 4553
 FT REPEAT 4554 4557
 FT REPEAT 4558 4561
 FT REPEAT 4562 4565
 FT REPEAT 4566 4569
 FT REPEAT 4570 4573
 FT REPEAT 4574 4577
 FT REPEAT 4578 4581
 FT REPEAT 4582 4585
 FT REPEAT 4586 4589
 FT REPEAT 4590 4593
 FT REPEAT 4594 4597
 FT REPEAT 4598 4601
 FT REPEAT 4602 4605
 FT REPEAT 4606 4609
 FT REPEAT 4610 4613
 FT REPEAT 4614 4617
 FT REPEAT 4618 4621
 FT REPEAT 4622 4625
 FT REPEAT 4626 4629
 FT REPEAT 4630 4633
 FT REPEAT 4634 4637
 FT REPEAT 4638 4641
 FT REPEAT 4642 4645
 FT REPEAT 4646 4649
 FT REPEAT 4650 4653
 FT REPEAT 4654 4657
 FT REPEAT 4658 4661
 FT REPEAT 4662 4665
 FT REPEAT 4666 4669
 FT REPEAT 4670 4673
 FT REPEAT 4674 4677
 FT REPEAT 4678 4681
 FT REPEAT 4682 4685
 FT REPEAT 4686 4689
 FT REPEAT 4690 4693
 FT REPEAT 4694 4697
 FT REPEAT 4698 4701
 FT REPEAT 4702 4705
 FT REPEAT 4706 4709
 FT REPEAT 4710 4713
 FT REPEAT 4714 4717
 FT REPEAT 4718 4721
 FT REPEAT 4722 4725
 FT REPEAT 4726 4729
 FT REPEAT 4730 4733
 FT REPEAT 4734 4737
 FT REPEAT 4738 4741
 FT REPEAT 4742 4745
 FT REPEAT 4746 4749
 FT REPEAT 4750 4753
 FT REPEAT 4754 4757
 FT REPEAT 4758 4761
 FT REPEAT 4762 4765
 FT REPEAT 4766 4769
 FT REPEAT 4770 4773
 FT REPEAT 4774 4777
 FT REPEAT 4778 4781
 FT REPEAT 4782 4785
 FT REPEAT 4786 4789
 FT REPEAT 4790 4793
 FT REPEAT 4794 4797
 FT REPEAT 4798 4801
 FT REPEAT 4802 4805
 FT REPEAT 4806 4809
 FT REPEAT 4810 4813
 FT REPEAT 4814 4817
 FT REPEAT 4818 4821
 FT REPEAT 4822 4825
 FT REPEAT 4826 4829
 FT REPEAT 4830 4833
 FT REPEAT 4834 4837
 FT REPEAT 4838 4841
 FT REPEAT 4842 4845
 FT REPEAT 4846 4849
 FT REPEAT 4850 4853
 FT REPEAT 4854 4857
 FT REPEAT 4858 4861
 FT REPEAT 4862 4865
 FT REPEAT 4866 4869
 FT REPEAT 4870 4873
 FT REPEAT 4874 4877
 FT REPEAT 4878 4881
 FT REPEAT 4882 4885
 FT REPEAT 4886 4889
 FT REPEAT 4890 4893
 FT REPEAT 4894 4897
 FT REPEAT 4898 4901
 FT REPEAT 4902 4905
 FT REPEAT 4906 4909
 FT REPEAT 4910 4913
 FT REPEAT 4914 4917
 FT REPEAT 4918 4921
 FT REPEAT 4922 4925
 FT REPEAT 4926 4929
 FT REPEAT 4930 4933
 FT REPEAT 4934 4937
 FT REPEAT 4938 4941
 FT REPEAT 4942 4945
 FT REPEAT 4946 4949
 FT REPEAT 4950 4953
 FT REPEAT 4954 4957
 FT REPEAT 4958 4961
 FT REPEAT 4962 4965
 FT REPEAT 4966 4969
 FT REPEAT 4970 4973
 FT REPEAT 4974 4977
 FT REPEAT 4978 4981
 FT REPEAT 4982 4985
 FT REPEAT 4986 4989
 FT REPEAT 4990 4993
 FT REPEAT 4994 4997
 FT REPEAT 4998 5001
 FT REPEAT 5002 5005
 FT REPEAT 5006 5009
 FT REPEAT 5010 5013
 FT REPEAT 5014 5017
 FT REPEAT 5018 5021
 FT REPEAT 5022 5025
 FT REPEAT 5026 5029
 FT REPEAT 5030 5033
 FT REPEAT 5034 5037
 FT REPEAT 5038 5041
 FT REPEAT 5042 5045
 FT REPEAT 5046 5049
 FT REPEAT 5050 5053
 FT REPEAT 5054 5057
 FT REPEAT 5058 5061
 FT REPEAT 5062 5065
 FT REPEAT 5066 5069
 FT REPEAT 5070 5073
 FT REPEAT 5074 5077
 FT REPEAT 5078 5081
 FT REPEAT 5082 5085
 FT REPEAT 5086 5089
 FT REPEAT 5090 5093
 FT REPEAT 5094 5097
 FT REPEAT 5098 5101
 FT REPEAT 5102 5105
 FT REPEAT 5106 5109
 FT REPEAT 5110 5113
 FT REPEAT 5114 5117
 FT REPEAT 5118 5121
 FT REPEAT 5122 5125
 FT REPEAT 5126 5129
 FT REPEAT 5130 5133
 FT REPEAT 5134 5137
 FT REPEAT 5138 5141
 FT REPEAT 5142 5145
 FT REPEAT 5146 5149
 FT REPEAT 5150 5153
 FT REPEAT 5154 5157
 FT REPEAT 5158 5161
 FT REPEAT 5162 5165
 FT REPEAT 5166 5169
 FT REPEAT 5170 5173
 FT REPEAT 5174 5177
 FT REPEAT 5178 5181
 FT REPEAT 5182 5185
 FT REPEAT 5186 5189
 FT REPEAT 5190 5193
 FT REPEAT 5194 5197
 FT REPEAT 5198 5201
 FT REPEAT 5202 5205
 FT REPEAT 5206 5209
 FT REPEAT 5210 5213
 FT REPEAT 5214 5217
 FT REPEAT 5218 5221
 FT REPEAT 5222 5225
 FT REPEAT 5226 5229
 FT REPEAT 5230 5233
 FT REPEAT 5234 5237
 FT REPEAT 5238 5241
 FT REPEAT 5242 5245
 FT REPEAT 5246 5249
 FT REPEAT 5250 5253
 FT REPEAT 5254 5257
 FT REPEAT 5258 5261
 FT REPEAT 5262 5265
 FT REPEAT 5266 5269
 FT REPEAT 5270 5273
 FT REPEAT 5274 5277
 FT REPEAT 5278 5281
 FT REPEAT 5282 5285
 FT REPEAT 5286 5289
 FT REPEAT 5290 5293
 FT REPEAT 5294 5297
 FT REPEAT 5298 5301
 FT REPEAT 5302 5305
 FT REPEAT 5306 5309
 FT REPEAT 5310 5313
 FT REPEAT 5314 5317
 FT REPEAT 5318 5321
 FT REPEAT 5322 5325
 FT REPEAT 5326 5329
 FT REPEAT 5330 5333
 FT REPEAT 5334 5337
 FT REPEAT 5338 5341
 FT REPEAT 5342 5345
 FT REPEAT 5346 5349
 FT REPEAT 5350 5353
 FT REPEAT 5354 5357

Db 1980 ALEILAHDEIDTKQ-----KUSSKLEIENGQIYLVKVLINQYKKNKISSIKSKEEAVSKI 2035
QY 923 THSK--EISNI--KDIIFNVNGLVKKVKNVDAHEVNTINSAFFIQ-----SLI 969
Db 2036 GNVSKHSELSKITCDKSYDNI-----IALEKQTELOQLNRSFTQCKTNTNSDSKLE 2088
QY 970 EYNTYKESLSNLSVANK-----VQVYAOI.FSTGLNTITDASKVELVSTA 1014
Db 2089 KIKTDFESLKNALKTLEGEVNALKASSDNHEVQSKSEBPVPAUSEIEKEBETDIDSUNTA 2148
QY 1015 LDETI-----DLLPTLSEGLPIIATIIDGVSLGAA-----IKELS 1049
Db 2149 LDELKKGKTCVSRKLIKOTVTKEISDDTELTINIEKNVAYLAYIKNYEDTVQDVL 2208
QY 1050 ETNDPLLRQIE-----AKIGIMAVNLTA--STAIVTSALGI----- 1085
Db 2209 TLINEHNTQVSNHEFTNFKNSKSEELTKAVTDSKTIISKLGVIENVNEMENTMTIE 2268
QY 1086 --ASGESILLVPLAGISAGISPL--VNNELIIOQ---KATKVIDYFKHISLAETGATL 1138
Db 2269 SGAKEIEALYNELKNKTKSLNIEYQTSNEVKLQEMKSNADKIIDVSKIFN-----TV 2320
QY 1139 LDDKIIMPQDDVLSEIDENNSITL-GRKEIWRAGGGSGHTLTDIDHFFSSPSITVRK 1197
Db 2321 LPTQ-----KSNIVTNQHSINNVRDKLKGKQ-----ELIDAGSFTLESI---K 2362
QY 1198 PWSIYDVL--NI-----KKEKIDFSDKMLVLPNAPRVGVEMGWPFGFRSLDN 1245
Db 2363 KFEIYSHIKTNLIGELOEQOQTNKSEHNDVAKHEKIVHLINRV-----ESLKG 2411
QY 1246 DGTQLLDRIRDHVGOYVRYFAP-----IADALITKLPREYDNTVRINLD 1292
Db 2412 D-----VKNHDDQYMKKLNASLLNDNIKNNTSINISDEELKLLKKVEE--NQQLCKN 2463
QY 1293 GNTRSITPVI--TTEQIRKNLSVSGSGGSVLSLSPN-----WNIDNLVENDTWID 1347
Db 2464 NNTQNFISDMKRVEDLNRFTENLPEKEKHQIE--NNYNEISSIFSEINLQDVDFVAK 2522
QY 1348 VDNVKNITTESDEIQGELIENILSKLNDENKILNHNHTNFYGDINESNRFTSLAPS 1407
Db 2523 IHKQIDAELKASVNNVREAKIRTAIONVTSYOTELL-----S 2559
QY 1408 ILSDINIIE-----IDLVSQY---KILLSGCMKLIENSSDIOQKID-HIGPNGE 1455
Db 2560 RLSEMNVLERTITRTKMDQLKSLSPDNTSLNARHTVRKSEDIILQNLASHIGEITE 2619
QY 1456 HQYI--PYSYDN-----ETKINGFIDYSKKEGLFTAEFNSIIRNI 1497
Db 2620 LNTYAHEVMYLENNELNKLKQLEIBRAKLETKTSPSGMKAKEEVPKPEKENRA----- 2674
QY 1498 YMPDSNNLFIYSSKDLKDIRINKGDVLLIGNYPKDDMKVLSFTIEDNTIKLGVYL 1557
Db 2675 ---QDNLASVPQKTLDE-----NTQOMPENSVDQNLASVPQKTLDE--NTHOMPENRV 2722
QY 1558 DENGVA 1563
Db 2723 QEDSIS 2728

RESULT 4
RBPI_PLAVB STANDARD; PRT; 2869 AA.
ID RBPI_PLAVB AC Q00798;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Reticulocyte binding protein 1 precursor.
GN RBPI.
OS Plasmodium vivax (strain Belen).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_taxID=31273;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites";
RL Cell 69:1213-1226(1992).
CC -I- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to
CC human reticulocyte cells.
CC -I- SUBUNIT: Homodimer (Potential).
CC -I- SUBCELLULAR LOCATION: Membrane-bound.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M88097; AAA29743.1; -
KW Malaria; Receptor; Signal; Transmembrane.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
FT DOMAIN 18 2807 EXTRACELLULAR.
FT TRANSMEM 2808 2826 POTENTIAL.
FT DOMAIN 2827 2869 CYTOPLASMIC.
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205BECFF CRC64;
Query Match 4.1%; Score 360; DB 1; Length 2869;
Best Local Similarity 19.6%; Pred. No. 3e-06;
Matches 386; Conservative 325; Mismatches 635; Indels 624; Gaps 103;
QY 6 KAOIQKVVYKFRIOEDEY-----VAIINALEEYHNMSVSVVEKYKLLK-DINNLL- 55
Db 946 KALKEKIVSDLRKIDQYETEFKEKTSAVENTVSTIQSLSKA-IDSLKRLANGSINNCK 1003
QY 56 ---TD-NYLTNYKSGGNKALK-----KPEYLTWEVLEKNSLTPVEKHLHFWIG 104
Db 1004 KYNTDIDLLASKIKTLREEVQKMPKRGDKCGENTTALLKSLDRKMGKINEKLN----D 1059
QY 105 QGIN--DT-----AINVINOWKDVN----- 122
Db 1060 GLNSLDTKEDLLKFYSKSKIHLSKQKQDPDLNRIDEDWDIKRDVDLNNVQVI 1119
QY 123 SDYTVKVFYDSNAPLNTLTKTI-----VESATNTLTESFRE-----NLNDPEFDYK 171
Db 1120 SENKVTIFKNSVTYIEAMSHINTVAHGITSNKNEILKSVKEVDKLVQNEQEDYK 1179
QY 172 ---YRKREMEI---YDKQKHFDYKSGIENPEFIDNIITYLSN-----EYSKOLEA 220
Db 1180 KNPENKQLEAIRGSMKLEKVINHVSEMTQ-----LESTANTLKSNAKGENEHDL 1234
QY 221 LNKY-----IEBSLKI-----TANGNDINLEKFADEDLVRLYNQELVERWNLAA 267
Db 1235 LNKTKGQMRDIYEKLEKIAEELKEGTYNELKDNKANKVEPPEPNIIGHVLE----- 1288
QY 268 ASDILRISMLKEDGVLDVLDLPGIQDLFKSINKPDSITNTSWEMKLEAINMYKEYI 327
Db 1289 ---RITVEKXKAGV--VEEMNSLTKIEKLEQ-----TSSDSQNELVTTSITKHL 1338
QY 328 PGYTSKNFDMLEDEEVQSFESALSSKSDSEIFFLPDDIDKVPSPLEVKI-----AFANNSVI 383
Db 1339 KGY-----EDVIKNEEDSIQIR-EKAKSLFTDEMCKLVQOVNMNLSQAQGNAGI 1389
QY 384 NQALISLKDYSCLDVLINQIRYKILNDNINPINEGTDFNTMKIFSDKLA-----S 437
Db 1390 SKEL-----NELKGVIELLISTYSSILEYVKKNSSESVRFSQLANGEFTKA 1436
QY 438 ISNEDNMFMFKITNYLKVGFAPDVRSTINLSGGVYTGAYQDILLMFKDNSTNIHLLEPE 497
Db 1437 EGEEKNAGARLAEAEKLEQIVKOL-----DYSN-----IDDK 1469

13-AUG-1987 (Rel. 05, Created)
01-MAR-1989 (Rel. 10, Last sequence update)
15-MAR-2004 (Rel. 43, Last annotation update)
DE Merzoite surface protein 1 precursor (Merzoite surface antigens)
DE (PMWSA) (P195).
GN MSP-1.
OS Plasmodium falciparum (isolate Camp / Malaysia).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5835;
RN [1]
RP SEQUENCE OF 1-1103 FROM N.A.
RX MEDLINE=86205236; PubMed=3517809;
RA Weber J.L., Leininger W.M., Lyon J.A.;
RT "Variation in the gene encoding a major merozoite surface antigen of
the human malaria parasite Plasmodium falciparum.";
RL Nucleic Acids Res. 14:3311-3323(1986).
RN [2]
RP SEQUENCE OF 1104-1726 FROM N.A.
RX MEDLINE=89143999; PubMed=3278296;
RA Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;
RT "Merozoite surface protein sequence from the Camp strain of the human
malaria parasite Plasmodium falciparum.";
RL Nucleic Acids Res. 16:1206-1206(1988).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
(Potential).
CC -!- PTM: Merozoite surface antigen contain the sequence of 83 kDa, 42
kDa and 19 kDa antigens which are the major surface antigens of
merozoites. The maturation take place during schizont.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03831; CRA27446.1; -.
DR PIR; A23386; SAZQGM.
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1726 MERZOITE SURFACE PROTEIN 1.
FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1726 AA; 196197 MW; D8RAD45FA352BCF3 CRC64;
Query Match 3.6%; Score 311.5; DB 1; Length 1726;
Best Local Similarity 19.8%; Pred. No. 0.00011;
Matches 349; Conservative 282; Mismatches 604; Indels 525; Gaps 89;
QY 51 DINNLTNDYNTYKSGRNKALKRFEYLTWEVLELKNLSITPVKKNLHFIWGGQINDT 110
Db 154 DLKRVNRYLT-----IKELKYPELPDLTNHMLTLC-D-NIH----- 189
QY 111 AINY-INQWQVNSYTVKVFYDSNAFLNTLTKTIVESATNNTLE-SFRENLDPEFDY 168
Db 190 GPKYLIDYEIN-----ELLYKLN-FYFDLLRAKLVANDVCANQCOIPFNLKTRANELDV 243
QY 169 NK----PYRKGMELIYDKQHFIDYKSGIENPEFIIDNIKIYLSNEYSKDLEALNKY 224

Db 244 LKLVFGYKPLDNKIKNVGKMEYIKN-----KTTIAN-----INEL 282
QY 225 IEESLNKITANGNDIRNLEK---FADEDLVRLYNQBELVERWNLAAASDILRSMLEKDG 281
Db 283 IEGSKKTIDQKNADNEEGKKLYQAQVDL-SYKQLEBAHNLISVLE-KRIDTLKNE 340
QY 282 GYILDVILPOIQ-PDLFKSINKPDSITNTSWEMIKLEAMKMYKEYIPGTYKKNFMDLDE 340
Db 341 NIKELLDKINEIKNPPPPANSNGTNTLLDKVKYIEEHEE--KIKE-IAKTIKFNIDSFT 397
QY 341 EVQSFESALSSKSDKBEIFPLDDIKVSPLEVIAFANNSVINQALISLMDSCSCLVI 400
Db 338 D-PIELEYLEKKNKVDVTPKSQDPTKSVQIPKVPYPNGIVPLPLTDIHNSLAAD--- 453
QY 401 NQIKNRY-KILNDNLNPSINEGTDFNTTKIPFSOKLASIGNEDNMFMKITYNLKVGFA 459
Db 454 -NDKNSYGLANPDTEKINE-----KIITD-----NKKIKIFI-----NNIK----- 490
QY 460 PDVSTINLSFGVGTGAYQDLMFKONSTWHLLEPELRNFEFPKTKISQITQEETISL 519
Db 491 -----KQIDLEEKKNHTKEQNKKLL 511
QY 520 WSNQARAKSQFEYKGYFEGALGEDDNLDFAGNTVLDDKYVSKKILSSMKTNKE--- 576
Db 512 EDYE-KSKDYEBELLEKFYE-----MKFNNN--FDKDVVD-KIFSARYTNVEKOR 558
QY 577 YIH-----YIVQLQCDKISYEASCNLPSKDPYSILYQKNIEGSETAYYYV---A 624
Db 559 YNNKSSNNNSVYNVQKLKALSY-----LEDYS--LRKGISSEKDFNHYTLTKGL 607
QY 625 DAIEIKDIKYPQVQISKNRN--IKLTIG-HGKSEPTDTFANLDV-DSLSSEIETIL 679
Db 608 EADIKKLT-----EIKSENKILEKFKGLTHSAN-----ASLEVYDVKLVQKQVL 655
QY 680 NLAKADIPKPIEINLLGNMFSYSISAEETYPGK-----LLLIKIDVSELMP-- 728
Db 656 LIKKIEDLRK-IEFLKNAQL-KDSIHVNTYKQNKPEPYLILVKKEVDKLEFIPKV 713
QY 729 -----SISODSITVSANQE-----VRINEEGKEIILDHSGKWNKESIIDK 771
Db 714 KMLKKEQAVLSITQPIVAASETTEDGHSHTLSQSGETEVTETE---TEETV--G 768
QY 772 ISSKEYISFNPKENKIIIVKSKYLHELSTLLOEIRNANSSDIDLEKVMLTECEINVASN 831
Db 769 HTTITVITILPPEKVKVENS-----IEHKSNDNSQALTKTVYKVKLDDEFTKS 816
QY 832 -----IDROIVEGRIEBAKQLTSDSINIK-----NEFKLESISDSL-Y 870
Db 817 YICHKYLIVSNSMDQKLELVNLTPEBENELKSCDPLDLLFNIGNNIPAWYSYDSMNN 876
QY 871 DLKHQGLDDSHFISFEDISKTENGFRIRFINKETGNSIFETKEI----- 917
Db 877 DLQH-----LFFELYQKEMTYLHLKKEENHKKLEEQKQITGTSTSGFNGNT 926
QY 918 --FSEVATHIS--KEISINIKTIFDN-----VNGKLKVKVNLDAAEVNTLNSAFFIQLI 969
Db 927 VNTASATHSNQONQSNASNTQNGVAVSGPVAVEESHDPPLTVLSINDLKGIVSLL 986
QY 970 EY-NTTK-ESLSNLSVAMKVQVYAOLFSTGLNTITD-----ASKVVE-LVSTALDET 1018
Db 987 NLGNKTKVPNPUTISTTEMEKFEYNILKNNDYFNDDIKQFVKNSKSVITGLTEQKAL 1046
QY 1019 IDLLPTLSEGLPIATIIDGVS-----GAAIKELSETNDPLRQIEBAK 1063
Db 1047 NDEIKKLDKTLQSLDFLYNKYKLDRLFNKKELGQDKMQKKLT-----LLKEQLSK 1101
QY 1064 IGIMAVNLTAATAITVTSALGIFASGSILL--VPLAGISAGIPSLVNNELIQDKATVI 1121
Db 1102 LNSL-----NNPHNVLQNFVFFKFKKEABEAEETENTLENTKILKHYKGLVK 1149
QY 1122 DYFKHISLAETEGATFLDDDKIIMPQDDL-----VLSEIDFN--NNSITLGKCEIWRAE 1173

Db 1150 YNGESSPLKT-----LSEVSIQTDNYANLEKFRVLSKIDGKLNLDNLHLGKKLSFLS 1203
Qy 1174 GSGHTLTDIDHFFSSPSITYRKPLWSIYDVNLNIKEKIDFSKDLMLVLPNANRVFGYE 1233
Db 1204 SGLHLITE-----LKEVI---KKNYKNTGNSP----- 1227
Qy 1234 MGWTPGFRSLDNGTKLLDRIDRHVEGQFYWRPARIADALITKL-----KPRVEDINVRI 1289
Db 1228 -----SENKKVNEALKSYEN-----FLPEAKVTVTVTPPODPVTPSPLSV 1268
Qy 1290 NLDGNTSRFIVPTTTEQIRKNSYSFYSGGSYLSLSLSPYNNMIDLNLV-----EN 1341
Db 1269 RVSSSGS-----TKETQIPTSGSL--TELQOVQOLQNYDEEDSLVLPFGSEDN 1321
Qy 1342 DTWV-----IDVNVVNKNTIIESDIQ-----KGLIENILS-KLINIED 1379
Db 1322 DEYLDQVVTGSAISVTMDNLSGPNENYDVLYLPLAGVYVSLKQKQKNIFFNLNLD 1381
Qy 1380 NKIILNHH--TINYPGIDNESN-----RFSILTSLEIDINIIEIDLVSYSKILLISGNC 1433
Db 1382 ---ILNRLKRRKYFLDVLSDLMQFKHISSEYIIED-----SFKLNSBQK 1426
Qy 1434 MKLTENSSDIOOKIDH--IGNGEHQKY---IPVSYIDNETKYNGFI----- 1475
Db 1427 NTLKSYKIKESVENDIKPAQEGISYEVKYLAKYKDDLESIKVKEEKEKFPSPPTT 1486
Qy 1476 -----DYSKKEGLFTAESNESIIRNIYMPDSNNLFYSSKDLKDIRINKGD----- 1523
Db 1487 PPSPAKTDEQKESKLPFLTN---IETLV-----NNL-----VNKIDDYLN 1526
Qy 1524 --VKLLIGNYFKDMKVSLSFTIEDTNTI--KLVGVYLDENGVQAQILKFWNNKASLNITS 1579
Db 1527 LKAKINDCNVEKDEAHVKIT-KLSDLKAIDDKID-LFKNHNDFEALKKLJND-----DTK 1579
Qy 1580 NSLMNPLESINIKNIFNNL 1599
Db 1580 KOMLGKLLSTGLVQNFNTI 1599

RESULT 7

ID MSP1_PLAF3 STANDARD; PRT; 1682 AA.
AC P19596; Q25921;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMWSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate ro-33 / Ghana).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5834;
[1]
RN SEQUENCE OF 1-1061 FROM N.A.
RP MEDLINE=88166657; PubMed=3327688;
RA Certa U., Rotman D., Matile H., Reber-Liske R.;
RT "A naturally occurring gene encoding the major surface antigen
RT precursor p190 of Plasmodium falciparum lacks tripeptide repeats.";
RL EMBO J. 6:4137-4142(1987).
[2]
RN SEQUENCE OF 1032-1682 FROM N.A.
RP MEDLINE=95354793; PubMed=7628566;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of
RT merozoite surface antigen-1.";
RL Exp. Parasitol. 81:47-54(1995).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -!- PTM: Merozoite surface antigen contain the sequence of 83 kDa, 42
CC kDa and 19 kDa antigens which are the major surface antigens of
CC merozoites. The maturation take place during schizont.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M35727; AAA29715.1; -;
CC EMBL; Y00087; CAA68280.1; -;
CC EMBL; Z35326; CAA84555.1; -;
CC InterPro; IPR006209; EGF_like.
CC Pfam; PF00008; EGF; 1.
CC Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1682 MEROZOITE SURFACE PROTEIN 1.
FT TRANSMEM 1666 1682 MEMBRANE ANCHOR (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 422 422 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 599 599 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 785 785 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 881 881 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 901 901 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 947 947 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1178 1178 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1569 1569 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1682 AA; 192462 MW; C82A1E159948CAD6 CRC64;
Query Match 3.6%; Score 309.5; DB 1; Length 1682;
Best Local Similarity 19.5%; Pred. No. 0.00013;
Matches 333; Conservative 279; Mismatches 587; Indels 513; Gaps 83;
Qy 36 NMSESSVVEKVLKLDINNLTNDVNTYKSGRNKALKKFKYELTMEVLEKNNLSLTPVE 95
Db 100 NPSDDSDSDAKSYADLKHRYQNYLFT-----IKELYPELFDITNHLMLTCD 147
Qy 96 KNLHFIWIGQINDTAINY-INQKDVNSDYTVKVFYDSNAFLINTLAKTTIVESATNTLT 154
Db 148 -NIH-----GPKYLDGVEIN-----ELLYKLN-FYDLELAKLNDVCANDYC 189
Qy 155 E-SFRENLDPEFDYNNK----FYKRMBEIYDKQKHFDIYKQSIQIENPEFIIDNIKTY 209
Db 190 QIPFNLKIRANELDVLKLVGRKPLDFIKDNGVKMEDYIKKN-----KTT 236
Qy 210 LSNYSVKOLEALNKYIBESLNKITANNNDIRNLEK----FADEILVRLYNGELVERWMLA 266
Db 237 IAN-----INELIEGSKTIDQNKQADNEEGKKLYQAYDLF-IYNKLOEAHNLI 287
Qy 267 AASDILRISMLKEDGGVYLDVDILPGIOPDLFKINKPDSITNTSWEMIKLEAIMKYKEY 326
Db 288 SVLE-KRIDTLKKNENIKKLEDDIDKIDAEKPTTGVNQILS-----LRLEKSRHEEK 341
Qy 327 IP-GYTSK-NFDMLEDEEVQSFESALSCKSEIFLPLDDDIKVSPLVEVIAFANNSVI 383
Db 342 IKEIAKTIKFNIDRLFTD-PLELEYLEKNNKVDVTPKSQDPTKSVQIPKVPFNGVLY 400
Qy 384 NQALISLKSDSYCSDLVINQIKNRY-KILNDNLNPSINEGTDFNTMTKIFSKLASIGNED 442
Db 401 PLPLTDIHNSLAAD-----NDKNSYGDLMVPTKEKINE-----KIITD-----NKE 442
Qy 443 NMFMKIKTNVLKVGFPADVSTNLSGPGVYTGAYQDLLMFKDNSTNIHLLEPELRNFE 502
Db 443 KRIFI-----NNIKKQIDL-----EENKIN 462
Qy 503 PPKTKISOLTEQETISLWSFNOARAKSOFEEVKYKYPFGALGEDDNLDFAQNTVLDDKYV 562
Db 463 HTKEQNKLLLE-----DYEKSKDYELLELEKFKYEMKNFNPNKND-----VVDKIFS 508
Qy 563 SKKILLSMKTR-NKEYIH-----YIVQGGDKISYEASCNLFSDPYGSILYQKNIEGSE 616
Db 509 ARITYNVEKQKNNKFSNNNSVNVQKLKALS------LEDYS-----LRKGISSEK 557

QY	617	TAYYYYV---ADABIKZIDKIRPIQYISNKN---IKUTFIG---HGKSEFNTDTPANLDV- 668
Db	558	FNHYTUKTGLEADIKKLE-----EIKSSKNKLEKFNFXGLTHSAN-----ASLEVS 605
QY	669	DSLSSEETITINLAKADISPKYIEINILGCMRPSYSISABEYPGK-----LLKI 719
Db	606	DIVKLQVQVLLIKKIEDLRK-IELFLKNAQL-KDSIHVNIYQPNKPPYVILVUKKE 663
QY	720	KDRVSELMP-----SISODSITVSANOYE-----VRINEBGRKREILDHSGK 760
Db	664	VDLKKERIPKVKMDLKKKEQAVLSITQPLVAASETTEDGGHSTHTLSQSGETEVTBTEE 723
QY	761	WINKESIIKDISKXYISNPKNKIKIVSKYIHELSTLLOETRNNANSSDIDLEKKVM 820
Db	724	TVGHITVT-----ITLPPKEVKVENS-----IEHKSNDNSQALTKTVY 763
QY	821	LTECEINVAAN-----IDRQVEGRIEBAKNLTS-----DSNIYKNEFKL 861
Db	764	LKLDLEFLTSYICHKXVILVNSNMDQKLEVNLTPEENELKSCDRDLDFLNQNNIPA 823
QY	862	IESISDSL-YDLKHQGLDDSHFISFEDISKTENGFRIRFINKEBTGNSIFIETEKEI--- 917
Db	824	MYSIYDSMNDLQH-----LPFELYQKEMYYLHLKKEENHIKKLEBPKQITGT 873
QY	918	-----PSEYATHIS--KEISNIXOTIFDN-----VNGKLVKKNLDAAEHVNTL 959
Db	874	STSSPONTTNTAQSAATHSNQOQNASSTNTQNGVAVSSGPAVVEESHDLPLTVLSIS 933
QY	960	NSAFFIOSLLEY-NTTK--EBSLNSLVAMK-----VOVYQAQFSTGLNT-ITDASKVVE 1009
Db	934	NDLKGIIVSLNLGNKTKVPNPLTISTEMEKFYENILKIMIPENDDIKQFVKSNSKVIIT 993
QY	1010	LVSTALDEITDLLPTLSEGLPIIATIDGVS-----GAAIKELSTND 1053
Db	994	GLTETQXNALNDEIKKLKDTLQLSFDLYNKYKLMDLFNKXKELGQDKMQIKILT--- 1049
QY	1054	PLRQETEAKEIGIMAVNLTAASTAIVTSALGIASGPSILL--VPLAGISAGISLVNNEL 1111
Db	1050	-LUKEQJESKNSJ-----NPNHVNQNFSVFNKKKBEAIAETENTLENTKI 1096
QY	1112	ILODKATKVIDYFKHISLAETGAFTLLDDKIMPODDL-----VLSEIDFN--NNSIT 1163
Db	1097	LLKHVKGVLKYNGESSPLXT-----LSEVSIQTEDNYANLEKFRVLKIDGLKNDNLH 1150
QY	1164	LQKCEIWRAGGGGHTLTDIDHFFSPSITYRKPMLSIYDVNLNIKEKIDPFKOLMVL 1223
Db	1151	LQKKKLSFLSGLHLITE-----LXEV1-----KXKNYTG 1181
QY	1224	NAPNRVGYEMGWTPGFRSLDNDGTKLLDIRDHVEQFYWRYPARTADALIKL---K 1279
Db	1182	NSP-----SENKKVNBAKSYEN-----FUPEAKVTVTVPQ 1215
QY	1280	PRYEDTNRINLDGNTSRFIVPVTITTEQIRKNLSYFSGGYSLSLSPYNNMIDNLV 1339
Db	1216	PDVTPSPLSVRVSSSGS-----TKETQIPTSGLL-TELQVQVQLQNYDEEDSLV 1268
QY	1340	-----ENDTWV-----IDVDNVVNKNTIESDIO-----KGELIEN 1370
Db	1269	LPFGESEDNDXYLDQVVTGEAISVTMDNLISGFENEYDVIYLPKLAGVYRSLKKQIEKN 1328
QY	1371	ILS-KLIENIKIILNNH--TINPYGDINEN---RFSILTSILEDINIIEIDLVS 1423
Db	1329	IFFNLNLD---ILNRLKRGKRYFLDVLESDDLQFKHISNEVIEED----- 1373
QY	1424	SYKILLSGNCWKLIENSDDIOQKIDH-IGFNGEHQKY---IPXSYIDNETKYNGFI--- 1475
Db	1374	SFKLLNSEQXNTLLKSYKVIKESVENDIKFAQECISYEVKLVAKYKODLESIKKVIKEEK 1433
QY	1476	-----DYSKKEGLTAEFNSNESIIRNIYMPDSNNLIIFYSKOLKDIRII 1519
Db	1434	EPSPSPPTTTPSPAKTDEQKESKFLPLTN---IETLY---NNL-----V 1473

Qy	1520	NGKDVKLLIGNVFQDKDMKVSFTIEDNTWK	1551
Db	1474	NK-----IDYTLNLKAKINDCNVEK	1494
		-----	-----
RESULT 8			
SPOT_SCHPO		STANDARD;	PRT; 1957 AA.
ID	SPOT_SCHPO	STANDARD;	PRT; 1957 AA.
AC	Q10411; Q9USE9;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	10-OCT-2003 (Rel. 42, Last annotation update)		
DE	Sporeulation-specific protein 15.		
GN	SP015 OR SPACIF3.06C.		
OS	Schizosaccharomyces pombe (Fission yeast).		
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;		
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;		
OC	Schizosaccharomycetes.		
NCBI_TaxID	4896;		
FN	NCBI_TaxID=4896;		
[1]			
RP	SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.		
RX	MEDLINE=20107136; PubMed=10639340;		
RA	Ikemoto S., Nakamura T., Kubo M., Shimoda C.;		
RT	"S. pombe sporulation-specific coiled-coil protein Spo15p is localized		
RL	to the spindle pole body and essential for its modification.";		
RL	J. Cell Sci. 113:545-554 (2000).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAN=972;		
RX	MEDLINE=21848401; PubMed=11859360;		
RA	Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,		
RA	Spours J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,		
RA	Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,		
RA	Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,		
RA	Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,		
RA	Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,		
RA	Jones K., Jones L., Jones M., Leather S., McDonald S., McLean J.,		
RA	Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,		
RA	Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,		
RA	Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,		
RA	Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,		
RA	Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,		
RA	Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,		
RA	Weldjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,		
RA	Gabel C., Fuchs M., Fritz C., Helzer E., Moestl D., Hilbert H.,		
RA	Borzyn K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,		
RA	Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,		
RA	Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Mottier S.,		
RA	Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,		
RA	Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,		
RA	Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,		
RA	Domiguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,		
RA	Chiruttli L., Lowe T., McCombie W.K., Paulsen I., Potashkin J.,		
RA	Spirakovski G.V., Ussery D., Barrell B.G., Nurse P.;		
RL	"The genome sequence of Schizosaccharomyces pombe.";		
RL	Nature 415:871-880 (2002).		
[3]			
RP	SEQUENCE OF 705-871 FROM N.A.		
RC	STRAN=968 h90;		
RX	MEDLINE=20423868; PubMed=10759889;		
RA	Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,		
RA	Hiraoka Y.;		
RT	"Large-scale screening of intracellular protein localization in living		
RT	fission yeast cells by the use of a GFP-fusion genomic DNA library.";		
RL	Genes Cells 5:169-190 (2000).		
CC	-!- FUNCTION: Has a role in the initiation of spore membrane		
CC	formation.		
CC	-!- SUBUNIT: Monomer.		
CC	-!- SUBCELLULAR LOCATION: Spindle pole body.		
CC	-!- SIMILARITY: Belongs to the MPC70 family.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collabora		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstati		

QY 1128 -SLAETGAP-TLLDDKIIPQDDVLSEID-FNNISITLCKEIRWAEAGSGHTLTDDI 1184
DQ 921 LSLGKNKIYQELTGQSSNFYKILKDSDTFYNESFT-----NFKSKADDINSLNDES 976
QY 1185 DHPSSPSITVRKPLSI-YDLNLIKKEID--FS-----KDLMLVLP----- 1223
DQ 977 KRKLEBEDINKKTLQSLFDLNYKYLKLERFDKKTKVGYKMQIKLTLKEQLESK 1036
QY 1224 ----NAPRVF-----GYEMGWTGFRSLDNDGTGKLDRIDHYEG----- 1260
DQ 1037 LNSLNNPKHVLQNFVSFFNKKBAEIAETE--NTLEN--FKILLK--HYGLVKYNGE 1089
QY 1261 ----QYWKYFAFIADALITKLPYEDTNVIRINDGNTRSHVP-----VI 1303
DQ 1090 SSPLKTLSESIQEDNYASLENPKVLSKLEGLKO--NLNLEKKLSYLSGLHLLA 1146
QY 1304 TTEQIRKNSLVSFGSGGSYLSLSPYNNIDNL-----1338
DQ 1147 ELKEVIKKNY-----GNPSSENNDVNNALESYKFLPECTOVATVVSBSGSD 1196
QY 1339 ----VENDTWI--DVDNVVKNIT-----ESDE-----IQKE-- 1366
DQ 1197 TLEQSPKPKASTHVGASNTITTSQNVDDVDVIVPIFGSEEDYDGLQVVTGEAV 1256
QY 1367 ----LIENILSKNTE-----DNKIILNNHTINFYGDINESNRFLSILFSL 1409
DQ 1257 TPSVIDNLSKIENEYEVLYKPLAGVYRSLKQLENNVTFNVVKD---ILNSRFNKR 1313
QY 1410 EDINIIEIDLVSYSKILSGNCM-----KLIENSSDIQKID-HIGFNG 1454
DQ 1314 ENFKNVLESLLI--PYKDLTSSNVVVKDPYKFLNKEKRFLESYNYKIDSIDTDNPNAN 1371
QY 1455 EQKVPYSYIDNETKNGRFDYSKGLFTAEFNSIIRNYMPSNNL-FIYSS-KD 1512
DQ 1372 D--VLGVYKILS-KYKSDLD-SIKKYINDKQENK-----YLPPLNNIETLYKTVD 1421
QY 1513 LKDIRIINKGVKLLIGNYKDDMKVSL-----FTIEDNTIKLNGVYLDEN--GVAQI 1565
DQ 1422 KIDLFVHL-EAKVLYNYTEKSNVEVKIKELNYLKTQD---KLADPKQNNFVGIADL 1476
QY 1566 LKFMNKAASLNTSNLNFLESNI-----KNIFYNNLDPNTEFILD 1608
DQ 1477 -----STDYNNHLLTKFLSTGWFENLAKTVLSNLLDGNLQGMN 1517
RESULT 10
MSPI_PLAFF STANDARD; PRT; 1701 AA.
AC P13819;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMWAS).
GN MSP-1
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5837;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88142999; PubMed=2449612;
RA Peterson M.G., Coppel R.L., McInyre P., Langford C.J., Woodrow G.,
RA Brown G.V., Anders R.F., Kemp D.J.;
RT "Variation in the precursor to the major merozoite surface antigens
RT of Plasmodium falciparum";
RL Mol. Biochem. Parasitol. 27:291-302(1988).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -!- PTM: Merozoite surface antigen contain the sequence of 83 kDa, 42
CC kDa and 19 kDa antigens which are the major surface antigens of
CC merozoites. The maturation take place during schizont.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; M19143; AAA29653.1; -;
DR PIR; A54498; A54498.
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 607 607 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 899 899 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 919 919 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1701 AA; 193719 MW; 3920875E73D38552 CRC64;
Query Match 3.5%; Score 307.5; DB 1; Length 1701;
Best Local Similarity 19.5%; Pred. No. 0.00015;
Matches 348; Conservative 287; Mismatches 615; Indels 533; Gaps 88;
QY 36 NMSESSVVEKYLKLDINLDVINTYKSGRKAALKKFKYLTMEVLEKNSLTPE 95
DQ 106 NPSDNSSSTKTADLKHRYQNYLT-----IKELYPFLDTNHLT-LS 152
QY 96 KNLHFIWIGQINDPAINQWQVNSDYTVKYFDSNAFLINTLTKTIVESTNNTLE 155
DQ 153 KNV-----DGFKYLDIGYBEIN-----ELLYKLN-FYDILLRAKANDACANSYCQ 196
QY 156 -SFRENLDNPPFDYNNK---FYRKMEIYDKQKHFDIYKASQIEENPEFIDNIKTYL 210
DQ 197 .PPNLKIRANELDVLKTVFGYRKLNDKNGVGMEDYIKKN-----KTTI 243
QY 211 SNEYSKOLEALKNYTEESLNKITANNNDIRNLEK--FADEDLVRNLQELVERWNLAAA 268
DQ 244 AN-----INELTEGSKTIDQKNADNEEGKKLYQAQYNYLFYVKNQLEAHNLISV 295
QY 269 SDILRISMLKEDGGVLDVLDLPGIOPDLFKSI--NKPDSITNTSWEMIKLEALMKYKEY 326
DQ 296 LE-KRIDTKKNENIKKLEIDIKITDAENPTTGSKPNLPENK-----KKE 342
QY 327 IPGYTSK-----NFDMLDEVRQVSFESALSSKSDKEIFLPLDDIKVSPLEVKIA 376
DQ 343 VEGHEEIKETAKTIKFNIDSFLT-D-PLEEYLYREKNKKVDVTPKSQDPTKSVQIPKVP 401
QY 377 FANNSEVINQALISLMSYCSDLVINQIKNRY-KIINDNLNPSINEGDTFNTTKIFSDKL 435
DQ 402 YFNGIVYPLPLTDIHNSLAAD---NDKNSYGLDMNPDTKEKINE-----KIITD-- 447
QY 436 ASINEDNNMFMKITNYLKVGFAPDVVSTNLSPGVYTGAYQDLMFKNSNTNHLLE 495
DQ 448 ----NKKRKIFI-----NNIKKQIDL-----464
QY 496 PELRNFPEPKYKISLTQELTSLWSFNQARAKSQFEYKKGYPEGALGEDDNLDPQONT 555
DQ 465 -BEKNINHTKQNKKLE-----DYKSKKDYELLEKFEY-----MKFNKN- 505
QY 556 VLDDKDVSKILLSSMKTRNKE---YIH-----YIVLQGDKISYEASCNLPSKDPY 603

Db 506 -PDKDWD-KIFSARYTYNVEKQRYNNKFSNNNSVNVQKXKALSY-----LEDY 555
QY 604 SLLYQKNIEGSEYAYVYV---ADAEIKEDIKYRIPYQISNKEN--IKLTFIG--HGKS 656
Db 556 S---LRKIGSEKDFNHYTYLTGTGEADIKKUTE-----EIKSSNKILEKRFGLTHSAN 607
QY 657 EFNTDTFANLDV-DLSISEIETILNLAKADISPKYIEINLLGCMFYSISAEETYPGK- 714
Db 608 ---ASLEVSIDVILQVQKVLIIKKIEDLRK-IELFLAKAQL-KDSIHVPNIYKPN 658
QY 715 -----LLKIKDRVSELMPISQ-----DSIT---VSANQVEVRINEEG--KR 752
Db 659 KPEPYLLVLKVEVDKLEFIPKVKMDMLKQBAVLSSITQPLVAASE-----TTEDGGHST 714
QY 753 EILDHSGKWINKERSIID-----ISSKEYISFNPKNKILVKSXYLHELSTL 800
Db 715 HTLSQSGTEVEETVEETVGTHTTITLPPKEESA--PKVKKVENS----- 763
QY 801 LOEIRNNANSDDLEKKVMLTECINVASN-----IDRQIVEGRIEAKNLTDSI 852
Db 764 ----IEHKNDSNQALTKVYIKLDEFILTKSYICHYILVNSNSMDQKLEVYNLTPEE 820
QY 853 NYIKNEFKL-----IESISDSLYDKHQGLDSDHFSIDISKTENGFRIRFINKETGN 907
Db 821 NELKSCDPLDLFNQNNIPAMYSLYDSMNIDLQHL--FFELYQKEMYIYLLHKLKEENHI 878
QY 908 SIFITEKEI-----FSEYATHIS--KEISNIKOTIFDN-----VNGKLVK 946
Db 879 KKLBEQKQITGTSSTSPGNTVTNTAQSAHNSQONQASNTQNGVAVSSGPAV 938
QY 947 KYNLDAAEHVNTLSAFTQSLIEY-NTTK-ESLSNLSVAMKVQVYALFSTGLNTITD- 1003
Db 939 BESHDPVLVLSISNDLKGIVSLNNGNKTVPNPLTISTTEWEPFENILKNNDTYNDD 998
QY 1004 -----ASKVVE-LVSTALDETIDLLPTLSEGLPIATIIDGVSL----- 1041
Db 999 IKQFVKSNSKVITGTETQKNALNDEIKKIDTLQSLFDLYNKYKLDRLFKNKKELGQ 1058
QY 1042 -GAATKESETNDPLRLQIEAKIGIMAVNLTAASTAVTSALGIASGFSILL--VPLAG 1098
Db 1059 DKWQIKKLT-----LLKEQLESKLSL-----NNPHNVQLQSVFFNKKKEAE 1101
QY 1099 ISAGIPSLVNNELIILQDKATKVIDYFKHISLAETGATLLDDKILMQDDLL-----V 1151
Db 1102 IAEENTLNTKILKHYKGLVYKNGESSPLKT-----SEVSIOETDNYANLEKFA 1155
QY 1152 LSEIDFN-NNSTIGKEIWEAEGSGHETLDDIDHFFSSPITYRKWLSIYDVLNKK 1210
Db 1156 LSKIDGKLDNHLGKKLSLSSGLHLITE-----L 1189
QY 1211 EKIDFSKDLMLVLPNAPNRVFGYEMGWTGFRSLDNDGTKLLDRIRDHYEGOFYRYFAFI 1270
Db 1190 EVI---KKNYTGNSP-----SENNKYNKALSKYEN-----FL 1220
QY 1271 ADALITKL-----KPYEDNVRINDGNTRSFIVPITTEQIRKNLSYFVGSGGSYLS 1326
Db 1221 PEAKVTVTVPQDPTPSPLSVRVSGSGS-----TKEETQIPTGSLI--TELQVQV 1273
QY 1327 LSPYNNIDNLV-----ENDTWV-----IDVNVKNITIESELIQ----- 1363
Db 1274 LQNYDEEDSLVLPPIFGESDNDEYLDQVVTGEAISTMDNLSGFENEDYVILVKPLA 1333
QY 1364 -----KGLIENILS-KLNIENDKILNNH--TNFYGDINES--RPTSLTFSTILE 1410
Db 1334 GYVRSLLKQIEKNIIIPNLND---LNSRLKKRYFLVDVLESOLMQFKLSSNEYIIE 1390
QY 1411 DINIIEIDLVSQKYLISGNOMKLIENSDDIQKIDH-IGFNGEHQKY---IPYSYID 1466
Db 1391 D-----SFKLLNSEQNTLLSKYIKESVENDIKFAQEGISYEKVLAKYKD 1438
QY 1467 NETKNGFI-----DYSKKEGLFAEFSNESIIRNIMPDNNL 1505
Db 1439 DLESIRKVIKEKEKFPSPPTTPPSPAKTDEQKESKFLPLTN-----IETLY-----NNL 1491

QY 1506 FIYSSKDLKDIRIINKGD-----VKLLIGNYFDDMKVSLSFSTIEDNTI--KLANGVY 1556
Db 1492 -----VNKIDDYLINLAKINCDCNVEDEAHVKIT-KLSDLKAIDDKID-LF 1536
QY 1557 LBENGVAQILKPMNNAKSALNTSLSIMNPLBSINIKNIFYNNL 1599
Db 1537 KNTNDPEAIKKLIND-----DTKMDLKGKLLSTGLVQNFNPNTI 1574

RESULT 11

MSPI_PLAPP STANDARD; PRT; 1726 AA.
ID MSPI_PLAPP AC P50435;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens) (PMMSA) (GP195).
GN MSP-1.
OS Plasmodium falciparum (isolate Palo Alto / Uganda).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=57270;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89005525; PubMed=3049134;
RA Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,
RA Siddiqui W.A.;
RT "Plasmodium falciparum: gene structure and hydropathy profile of the
RT major merozoite surface antigen (gp195) of the Uganda-Palo Alto
RT isolate";
RL Exp. Parasitol. 67:1-11(1988).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (potential).
CC -!- PTM: Merozoite surface antigen contain the sequence of 83 kDa, 42
CC kDa and 19 kDa antigens which are the major surface antigens of
CC merozoites. The maturation take place during schizont.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M37213; AAA29611.1; .
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF_1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 133 133 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 272 272 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 567 567 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 638 638 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 827 827 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 924 924 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 944 944 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 990 990 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1016 1016 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1114 1114 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1221 1221 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1613 1613 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 1726 AA; 196174 MW; 5859CEFA2F9A026 CRC64;

Query Match

Best Local Similarity 19.8%; Score 306.5; DB 1; Length 1726;
Matches 349; Conservative 284; Mismatches 602; Indels 525; Gaps 90;

QY 51 DINNLTDNYLNTYKKGGRNKALKKFKLEYLTMEVLELKNNSLTPEKKNLHFIWGGQINDT 110

RX MEDLINE=86136024; PubMed=3004972;
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RA Stunnenberg H., Bujard H.;
RT "Polymorphism of the precursor for the major surface antigens of
RT Plasmodium falciparum merozoites: studies at the genetic level";
RL EMBO J. 4:3823-3829 (1985).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -!- PTM: Merozoite surface antigen contain the sequence of 83 kDa, 42
CC kDa and 19 kDa antigens which are the major surface antigens of
CC merozoites. The maturation take place during schizont.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X05624; CAA29112.1; -;
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19
FT CHAIN 20 1701
FT CARBOHYD 110 110
FT CARBOHYD 239 239
FT CARBOHYD 470 470
FT CARBOHYD 536 536
FT CARBOHYD 607 607
FT CARBOHYD 802 802
FT CARBOHYD 899 899
FT CARBOHYD 919 919
FT CARBOHYD 965 965
FT CARBOHYD 991 991
FT CARBOHYD 1089 1089
FT CARBOHYD 1196 1196
FT CARBOHYD 1588 1588
SQ SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AF96EA98 CRC64;

Query Match
Best Local Similarity 3.5%; Score 304.5; DB 1; Length 1701;
Matches 355; Conservative 284; Mismatches 603; Indels 549; Gaps 91;

QY 36 NMSESVVEKYLKLDINNDYNTYKSGRNKALKKPKFYLTMEVLEKNNSLPVE 95
DB 106 NPSDSSDNTKYADLKHRYQNYLFT-----IKELKYPELFDLTNHEMLT-LS 152

QY 96 KNLHFIWIGGQINDTAINYNQWVDNSDYTKVYFDSNAPLINTLKKTIVESATNTLE 155
DB 153 KNV-----DGFKYLDIGYEEIN-----ELLYKLN-FYDILLRAKLANDACANSVQC 196

QY 156 -SFRENLDPEFDYK----FYRKEVEIYDKQHFIDYKSOEENPEFIDNIKTYL 210
DB 197 IPFNLIKIRANEDLVLLKIVFGYKPELDNKNVGMEDYIKGN-----KTTI 243

QY 211 SNEYSKOLEALNKYTEESLNTKITANGNDIRNLEK--PADSDVRLVNLQELVERWNLAAA 268
DB 244 AN-----INELGSKTIIDQKNADNEGKKLQAOVNIPIYKQLQEAHNLISV 295

QY 269 SDILRISMLKEDGGYLDVDPGLPGIPLFKSI--NKPDSITNTSWEMIKLEA-IMKYKEY 326
DB 296 LE-KRIDTLKKNENIKULEEDIKIKTDAENPTTSGKPNLPENK-----KKE 342

QY 327 IGYTSK-----NFDMLDEEVQSFESALSCKSDKEIFLPLDDIKVSPLEVKIA 376
DB 343 VEGHEEKIEAKTIKFNIDSFLTQ-PLLEYLYLRKNKVDVTPKSDQPTKSVQIPKVP 401

QY 377 FANNVINGQALLSLKSDYLVIQNKRY-KIINDNLNPSINEGTFNTMTKIFSDKL 435
DB 402 YPNGIVYPLPLTDIHNLAAD-----NDKNSYGLMNPDTKEKINE-----KIITD-- 447

QY 436 ASISNEDNMFMKITYNLKVGFPADVSTINLSPGVYVYTGAYQDMLMFKDNSTNIHLE 495
DB 448 ----NKRKIPI-----NNIKQIDL----- 464

QY 496 PLRNFPEPKWISOLTEQETISLWSFNQARAKSOFEBYKGYEGALGEDNDLPAQNT 555
DB 465 -BEKINHTEQNKLE-----DYKSKDYEELEKFBY-----MKFNPN- 505

QY 556 VLDDKVVSKILLSSMKTRKE--YIH-----YIVLQGDKISYEASCNLFSKDPY 603
DB 506 -FDKVDV-KIFSARYTNNVEKQRNNKFNSSNNSVYVQKLKALS------LEDY 555

QY 604 SSILYQKNIQEGSETAYIVYV--ADAIEIKDKIRIPYQISNKRN--IKLPIG--HGKS 656
DB 556 S---LRKGISEKDPNHYITLKTGLEADIKLKE-----EIKSENKILEKQKFGKLTGSAN 607

QY 657 EFNTEFTANLDV-DSLSBEIETILNAKADISPKVIEINLLGCNMFVSISAEETYPQK- 714
DB 608 ----ASLEVSIVLQVQVLLIKKIEDLRK-IELFLKNAQL-KDSIHVPNTYKPN 658

QY 715 ----LLALKIDRVSELMPISIQ-----DSIT--VSANOYEVRINEEG--KR 752
DB 659 KPEPYVILVKKEVDKLKEFIPKVKMDLKKQAVLSSITQPLVAASE-----TTEDGGHST 714

QY 753 EILDHSGKWINKESIIKD-----ISSKEYISFNPKNKIIVKSKYLHELSTL 800
DB 715 HTLSQSGTEVTEETEVEETVGHVTTTITLPPKEESA--PKEVKVENS----- 763

QY 801 LQEIERNANSSDIDLEKVMYTEC-----INVA-SNIDROIQVEG--RIBEA 844
DB 764 ---IEHSNDNSQALTKTVLKKLDEFLTKSYICHKYLVSNSMDQLLEVNLTPSEE 820

QY 845 KNLTS-----DSINYKNEFKLIESDSL-YDLKHQNGLDSDHSPISFDISKTENGFRIR 899
DB 821 KELKSCDPLDLFNIGNNIPAMYSLYDSMNLDQH-----LFFELYQKEMYYLH 870

QY 900 FINKETGNSIFETEKEI-----FSEYATHIS--KEISNIXDITPDN--- 939
DB 871 KUKENHIKLEEQKQITGTSTSPGNTVTNTAQATHSNSQCSNASSTNTQNGVA 930

QY 940 -VNGKLVKVNLDAAHEVNTLNSAFFIQSLIBY-NTTK-ESLSNLSVAMKVQVYQLFT 996
DB 931 VSSGPAVVEESHDPLTVLSINDLKGIVLLNGLNKKTKVPNPLTISTTEMEKFENILKN 990

QY 997 GLNITD-----ASKVVE-LVSTALDETTDLPTLSEGLPIIATIIDGVS- 1041
DB 991 NDTYFNDIDIKQFVKSNSKVITGLTETQKNALNDEIKKLDLQLSFDLYNKYKLDRLF 1050

QY 1042 ----GAAIKELSETNDPLLRQEIHAKTIGIMAVNLTAATAVTSALGIASGFSIL 1092
DB 1051 NKKELGQDKMQIKKLT-----LLKQESKNSL-----NNPHNVLQNFVSF 1093

QY 1093 L--VPLAGISAGIPSLVNNELIQDKATKVIDYFKHISLAETEGAFITLDDKIIMPQDDL 1150
DB 1094 FNKKKEAEIAETENTLENTKILLKHVGLVKYNGESSPLKT-----LSEVISQTEDNY 1147

QY 1151 ----VLSEIDFN-NNSITLGKCEIWRAGEGSGHTLTDIDHFPSPSITIRKPWLSI 1202
DB 1148 ANLEKFRALSKIDGKLDNLDNLHLGKKLSFLSSGLHLLTE----- 1187

QY 1203 YDVLNTKKKIDFKIDKDLAVLPNAPNRVFGYMGWTPGFRSLNDGDKLLDRIDHYEGOF 1262
DB 1188 ----LKEVI---KNKNTGNSP-----SENNKKNKALKSYEN-- 1218

QY 1263 YNRYFAFIADALTKL-----KRYEDTVNRINLDGNTRSFIVPVITTEIQRKNLSVFPYG 1318
DB 1219 ----FFPEAKVTVTTPPQDPVTPSPUSRVVSSGS-----TKETQIPTSGLL- 1265

QY 1319 SGGSYSLSPYNMIDNLV-----ENDTW-----IDVNVVKNITIED 1360
DB 1266 TELQVQVQLQNYDEEDDSLVLPITGESEDNDEYLDQVVTGSAISVTMDNLSGFENEYD 1325

QY 1361 EQ-----KGLIENILS-KLNIEDNKIILNH--TINFYGDINESN-----RPI 1402
Db 1326 VYLKPLAGVYSLKQIEKNIITFNALND---TNSRLKRRKFLDVLSDLMQPKH 1382
QY 1403 SLTFSLEINIIIEIDLVSYSKILLSGNOMKLIENSDDIOQKIDH-IGENGHOKY-- 1459
Db 1383 SNEYIIE-----SFKLNSQKQTLKSKYIKESVENDIKFAQSGISYE 1430
QY 1460 -IPYSIDNETKNGFI-----DYSKEGLFTAESNESIIRNI 1497
Db 1431 KVLAKYKODLESKKVKEEKFPSSPTTPPSAKTDEQKESKFLPFLTN---IETL 1487
QY 1498 YPDSNNLFIYSGKOLKOTRIKGD-----VKLLIGNYKDKMKSLSFTIETDNTI 1550
Db 1488 Y-----NNL-----VNKIDYILINKAKINDCNVEKDEARVKIT-KLSDLKAI 1529
QY 1551 --KLNGVYLDENGVAQILKFMNNAKSALNTSLSLAKFILESINIKNIFVNNL 1599
Db 1530 DDKID-LFKNTNDFEAKKLIND-----DTKDKMLGKLLSTGLVQNFNTI 1574

RESULT 13
YCF2_WARPO STANDARD; PRT; 2136 AA.
AC P09975;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical 259.9 kDa protein ycf2 (ORF 2136).
GN YCF2.
OS Marchantia polymorpha (Liverwort).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Eukaryota; Viridiplantae; Marchantiidae; Marchantiales; Marchantiaceae;
OC Marchantiaceae; Marchantia.
OX NCBI_TaxID=3197;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohyama K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
RA Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,
RA Ozeki H.;
RT "Chloroplast gene organization deduced from complete sequence of
RT liverwort Marchantia polymorpha chloroplast DNA.";
RL Nature 322:572-574 (1986).
CC -!- SIMILARITY: Belongs to the ycf2 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X04465; CAA28078.1; --
DR PIR; S01591; A05037.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003599; AAA_ATPase_centri.
DR Pfam; PF00004; AAA; 1.
DR SMART; SM00382; AAA; 1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2136 AA; 259911 MW; 5BD170C9CCF61197 CRC64;

Query Match 3.5%; Score 302.5; DB 1; Length 2136;
Best Local Similarity 19.3%; Pred. No. 0.00032;
Matches 376; Conservative 282; Mismatches 593; Indels 695; Gaps 101;

QY 27 ILNALEYHNSSESVVEKYLKLDINNL-TDNYL-----NTYKSGRNKALKKPKYEYLT 80
Db 309 IFKQLONF-NESDKLIESFLLKIKGNLYFKNVEPVTWQSYKKDCLD--FNKNELNN 365
QY 81 MEV-LELKNLSLTVEKLNHFIWGGQINDTAI-----NYINQMKVNS--DYTVKVFYDS 133

Db 366 SEIYIKIEELFSDYIYKFSKYILYEGKSKTKIILQSFNNNIYKLNLSIFNFTIETFD 425
QY 134 NAPLINTLKKTIIVESATNTT--LESFRENLDPEFY-----NKFYKRMELIYDKOKHFI 187
Db 426 NLLFLDWLKKVY--INNKPFLKSF-----LIYSSISNQF-----LFFKQKNSK 468
QY 188 DYKYSQIEENPEFIIDNIKTYS--NEYSKDL-----BALNKYIEESLKIT 233
Db 469 SFKNLKKKSKDVITNVFSKENKINNFPSKSIYAFPEILSINIDNKFV--INKIS 525
QY 234 ANNGNDIRN---LEKFADEDLVRLNQELVERWN-----LAAADILIRIS-ML 277
Db 526 LKINKKKQKRFYLNKIKSSDNFNLWKIKYSSQQFVSNNSFLNPAFELQQNYL 585
QY 278 KPDGQVLDVILPGIQDPLFSINKPDS-----ITNWSMEMIKLEAIMK----- 322
Db 586 KKNILF-----FKKLNVEFSNFFYQVYKCKKILFLKFAKLEKILKGRNK 632
QY 323 -----YKEYTPGYTSKNFMDLDEVQSFSSALSSKSKSEIPLPLDDIKV----- 368
Db 633 KFTISIKLFFKFKYKNKLNENGEYKIESQILQNEKELNKKKQKQFNP--NIKILSFYNS 690
QY 369 SPLEVKTIA--FANNSVINQALISLSDSYGSLVINQIKURYKILNDNLNPSINEGTFDN 425
Db 691 SKNLYLQNKYFNKNLNNKLITWK-----KISNKLVISNENY----- 730
QY 426 TTKIFSDKLASINEDNMFMKIITNYLKVGPADVRSTINLSGFGVYTGAYQDLIMFX 485
Db 731 --KI-----IWNKQKMF-----FSPKSNVY-----DTFFFN 756
QY 486 DNSTNIHLPELARNPEFPPTKTSQTEQITSLMS-----FNOARAKSOFEBYKGYFE 540
Db 757 KKSFNITV-----IFDKLKKIQLNFQEIQILNCFSLFFNSKNIK-KTKIFNSYF- 807
QY 541 GALGEDNDLPAQNTVLDKYVSKILSSMKTKNEYTHYIVOLQGDKIYSEASCNLF 600
Db 808 -----INENLTTFSTFNDKEFNIFLELFISEINNDLMRPFKK 846
QY 601 DPYSSILYQKIEGSETAYVYVADABE--KEIDKVRIPYQISNKNIKLTFIGHGKSEF 658
Db 847 -----VLYRIYKDELLFPNPIENROLLQNPFEKTKI-LTFI-----DF 884
QY 659 NTDTFANLDVDSLSSEIETILNAKADISPKYIEINLLGCNMFPSYSISABETVPGKLLK 718
Db 885 LQD-----PELANNRFRIFLEKTKIKNN---NLL-----YLRLKIF 919
QY 719 IKRVSELMPSISODSTVSAQVVEINEEGKREILDHGKWKINKEESIKDISSEYI 778
Db 920 LKDKRNFLL-----INE-----IK----- 933
QY 779 SPNPKENKIIVKSK-----YIHELSTLLQETRNANSSDIDLEK 817
Db 934 SFIEKKNNLFIKSQLSNVLLVKNYSKFFDNIFNPHFLKQKKNIEILNNQNYFEKSLK 993
QY 818 KVMILTECEINVA-SNIDROIVEGRIEBAKSLTSDSINVYKNEFKLIESISDSLY-----D 871
Db 994 KTYLKNLNNNSYSKFKIF--IFOLLNIN-----KNYKTFQWISLIELIYSKNLN 1044
QY 872 LKHQGLDSDSHFISFEDISKTENGFR-IRFINK-----ETGNSIFETEKEIFSEYATHIS 926
Db 1045 YKIQNKIEKNYCNKNISYKKKIKTVNPFKNLFTNNSWFFTE--WWEYNTYIL 1101
QY 927 KEISNICKTIFDNVN-----GKLVKVKNLDAAEVNTLNSAPFIQS----- 967
Db 1102 LQI--IQETFFQITDVLVEYFKKKKIIEK-----NLKPFKSKISLKTLSFHN 1147
QY 968 -----LIEYNTTESLSN-----LSVAMKVQVYVQ-LFSUGL 998
Db 1148 FKLKNLRFNEINRYKNYLNFLWSPFNLNCCNLYWVLSVIFSLVIFLYQKIFSLI 1207
QY 999 NT-----ITDAS-----KVVELVSTALDEITDILPTLSEGLPIAT----- 1034
Db 1208 GSDCHLWKNFEIQLYTDTRSLSYFTKLTRENKALTALNTENLLSYFFQNLTHYITNKF 1267

QY 1035 -----IIGVSLGAIKELSETNPDLLROEI--PAKIGIMAVNLTAASTALVTS 1081
 DB 1268 YLLTKNLKWLNNKTL-----DLRRKRLVQSLITNKIQNGVFELN-SNKOFFTS 1321
 QY 1082 ALG-----IASGFSILLVPLAGISAGIPSLVNNELIILQDKATKVIDY--- 1123
 DB 1322 YFGIQTINQQLLVFVLAQFFQ-----KNLNNLSL---DLANKWIVSFWSH 1365
 QY 1124 --FKHISLAETEGAFLLDDKIMPODDLVLSEIDFNNSITL-----GKCEIWRAE 1173
 DB 1366 KIFFSQKLRLQNNI-----ELGFQNIIPVLPQGLSFGKILLIGPIE 1407
 QY 1174 GSGHSLTDDIDHFF-----SSPSITYRKP-----WLSI-YDVNLKKEKIDFSKOLM 1220
 DB 1408 TGRSYLIKNAAESYVPLFKISINKLLYNKPDVITESMMNLLIESRLRLNLTLDPAKMM- 1466
 QY 1221 VLPNAPRVFCYEWGWPFFSL-----DNDGTLKLDRIHDHVEGOF---YWRYPFA 1268
 DB 1467 ----SPECII-----WQNIHQNLNRLTONVESDPTFLGILLKYFQDTFKTKNNII 1516
 QY 1269 FIADALITK-----LKPVED--TNVRI-NLDGNTSRFIVPVITTE---QIRKNLSYSF 1316
 DB 1517 VIGSTHLPKVDPALISNRLDKIINVLFNLSQKKQF--PLLKKKNFQKRENL--PF 1572
 QY 1317 YSGSGSYLSLSPYNNMIDLNVENDTWV-----IDVDNV-----VKNITIEDS 1360
 DB 1573 LNEFGSRMG--YNLR-DLSALTNVLLISITKNGRSFIDTDTLKLAFHRQIFGLTYTNN 1628
 QY 1361 BIQKGLLENLTSKL--NIEDNKILANHTNFP--YGDINESNPFSLTFSILE---DIN 1413
 DB 1629 KINFDRIFKIVYKVGKTIION--ILIKSSMNLNIGNFLWKKNFYLSLWYLEPSIDES 1687
 QY 1414 IIEIDLVSKSVKIL--LSGCMK-----LIENSDDIQKIDHIGFNGEHQKYPIYSYDN 1467
 DB 1688 IIKETILT---HILACLAGTAARDSWELKEKKAESLLPIDKL---VENDFTLAFLSLES 1741
 QY 1468 -----ETKYNGFIDYSKKEGLFTAESNESIIRNIYMPDSNNLFIYSSK 1511
 DB 1742 FSEPPWLEICOTNVNSKKNKILIEFTKNSM-----NIMQNGIPAIAANKKFIYTON 1793
 QY 1512 DLKDIRIKNGVDKLLIGNYFKDDMKVSLST----- 1543
 DB 1794 HLQ-----YKSLSQISFNKKNKQVEFKNTSPPRFLRSLPFRSLNLF 1836
 QY 1544 -IEDNTYIKLVYLDENGVAQLKFMNAKSALNTSNLSNLESLNINIKNIFNNLDPN 1602
 DB 1837 WIKRPNDFEFSYKF---GFTKKGEYLFSA--LQKKNYQGFIEKKKKQLLYERILPR 1890
 QY 1603 I-----EFILDTNFIISG 1615
 DB 1891 IRRNVQELSESQFEILLIEEQFEILG 1916

RESULT 14
 ID UTRO HUMAN STANDARD; PRT; 3433 AA.
 AC P46339;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Utrophin (Dystrophin-related protein 1) (DRP1) (DRP).
 GN UTRN OR DMDL (Human).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93096045; PubMed=1461283;
 RA Tinsley J.M., Blake D.J., Roche A., Fairbrother U., Kiss J.,
 RA Byth B.C., Knight A.E., Kendrick-Jones J., Suthers G.K., Love D.R.,
 RA Edwards Y.H., Davies K.E.;

RT "Primary structure of dystrophin-related protein.";
 RN Nature 360:591-593 (1992).
 RP INTERACTION WITH SNTB1.
 RX MEDLINE=95146543; PubMed=7844150;
 RA Ahn A.H., Kunkel L.M.;
 RT "Syntrophin binds to an alternatively spliced exon of dystrophin.";
 RL J. Cell Biol. 128:363-371 (1995).
 RN [3]
 RP INTERACTION WITH SNTA1 AND SNTB2.
 RX MEDLINE=96162017; PubMed=8576247;
 RA Ahn A.H., Feener C.A., Gussone E., Yoshida M., Ozawa E., Kunkel L.M.;
 RT "The three human syntrophin genes are expressed in diverse tissues,
 RT have distinct chromosomal locations, and each bind to dystrophin and
 RT its relatives.";
 RL J. Biol. Chem. 271:2724-2730 (1996).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 144-261.
 RX MEDLINE=99141377; PubMed=9887274;
 RA Keep N.H., Norwood F.L.M., Moores C.A., Winder S.J.,
 RA Kendrick-Jones J.;
 RT "The 2.0-A structure of the second calponin homology domain from the
 RT actin-binding region of the dystrophin homologue utrophin.";
 RL J. Mol. Biol. 285:1257-1264 (1999).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF 31-256.
 RX MEDLINE=20113481; PubMed=10647184;
 RA Keep N.H., Winder S.J., Moores C.A., Walke S., Norwood F.L.M.,
 RA Kendrick-Jones J.;
 RT "Crystal structure of the actin-binding region of utrophin reveals a
 RT head-to-tail dimer.";
 RL Structure 7:1539-1546 (1999).
 CC -!- FUNCTION: May play a role in anchoring the cytoskeleton to the
 CC plasma membrane (By similarity to dystrophin).
 CC -!- SUBUNIT: Interacts with the syntrophins SNTA1; SNTB1 and SNTB2.
 CC -!- SUBCELLULAR LOCATION: Neuromuscular junction.
 CC -!- TISSUE SPECIFICITY: Muscle.
 CC -!- SIMILARITY: STRONG, TO DYSTROPHIN.
 CC -!- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY
 CC ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN,
 CC ABP-120, ABP-180, OR BETA-PODRIN).
 CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
 CC -!- SIMILARITY: Contains 1 WW domain.
 CC -!- SIMILARITY: Contains 20 spectrin repeats.
 CC -!- SIMILARITY: Contains 1 ZZ-type zinc finger.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X69086; CAA48829.1; -.
 DR PIR; S28381; S28381.
 DR PDB; 1BHD; 16-FEB-99.
 DR PDB; 1QAG; 01-JAN-00.
 DR Genew; HGNC:12635; UTRN.
 DR MIM; 128240; -.
 DR GO; GO:0005856; C:cytoskeleton; TAS.
 DR GO; GO:0005624; C:membrane fraction; TAS.
 DR GO; GO:0005886; C:plasma membrane; TAS.
 DR GO; GO:0006936; P:muscle contraction; TAS.
 DR GO; GO:0007517; P:muscle development; TAS.
 DR InterPro; IPR001589; Actbind actinin.
 DR InterPro; IPR001715; Calponin-like.
 DR InterPro; IPR002017; Spectrin.
 DR InterPro; IPR001202; WW_Rsp5_WWP.
 DR InterPro; IPR000433; ZnF_ZZ_
 DR Pfam; PF00307; CH; 2.
 DR Pfam; PF00435; spectrin; 18.
 DR Pfam; PF00397; WW; 1.

```

DR Pfam: PF00569; ZF: 1.
DR SMART: SM00033; CH: 2.
DR SMART: SM00150; SPEC: 19.
DR SMART: SM00456; WW: 1.
DR SMART: SM00291; ZF: 2.
DR PROSITE: PS00019; ACTININ_1; 1.
DR PROSITE: PS00020; ACTININ_2; 1.
DR PROSITE: PS00021; CH: 2.
DR PROSITE: PS01139; WW_DOMAIN_1; 1.
DR PROSITE: PS00020; WW_DOMAIN_2; 1.
DR PROSITE: PS01357; ZF: 2.
DR PROSITE: PS0135; ZF: 2.
DR Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;
KW Repeat; 3D-structure; Actin-binding; Zinc-finger.
FT DOMAIN 1 246 ACTIN-BINDING.
FT DOMAIN 31 135 CH 1.
FT DOMAIN 150 252 CH 2.
FT REPEAT 253 308 SPECTRIN 1.
FT REPEAT 309 417 SPECTRIN 2.
FT REPEAT 418 526 SPECTRIN 3.
FT REPEAT 541 637 SPECTRIN 4.
FT REPEAT 687 798 SPECTRIN 5.
FT REPEAT 803 902 SPECTRIN 6.
FT REPEAT 1016 1083 SPECTRIN 7.
FT REPEAT 1125 1230 SPECTRIN 8.
FT REPEAT 1248 1334 SPECTRIN 9.
FT REPEAT 1432 1541 SPECTRIN 10.
FT REPEAT 1544 1649 SPECTRIN 11.
FT REPEAT 1652 1753 SPECTRIN 12.
FT REPEAT 1910 1968 SPECTRIN 13.
FT REPEAT 1976 2081 SPECTRIN 14.
FT REPEAT 2258 2333 SPECTRIN 15.
FT REPEAT 2399 2440 SPECTRIN 16.
FT REPEAT 2443 2556 SPECTRIN 17.
FT REPEAT 2559 2636 SPECTRIN 18.
FT REPEAT 2658 2688 SPECTRIN 19.
FT REPEAT 2691 2797 SPECTRIN 20.
FT DOMAIN 2812 2845 WW.
FT ZN FING 3064 3111 ZZ-TYPE.
FT HELIX 32 46
FT TURN 47 49
FT TURN 56 62
FT HELIX 64 74
FT TURN 75 75
FT HELIX 86 102
FT TURN 103 104
FT HELIX 112 117
FT TURN 118 118
FT HELIX 120 134
FT TURN 135 135
FT HELIX 136 149
FT HELIX 152 163
FT TURN 164 164
FT TURN 168 169
FT HELIX 177 179
FT TURN 180 181
FT HELIX 183 191
FT TURN 192 192
FT HELIX 194 196
FT HELIX 199 204
FT HELIX 207 222
FT HELIX 230 233
FT HELIX 240 253
FT TURN 253 253
SQ SEQUENCE 3433 AA; 39488 MW; EAB8DB409F858E5B CRC64;

Query Match 3.38; Score 289; DB 1; Length 3433;
Best Local Similarity 19.6%; Pred. No. 0.0019;
Matches 318; Conservative 293; Mismatches 612; Indels 488; Gaps 85;

QY 27 ILNALLEEYHNMSSESSVEKYLKLDINNLTDTNLTNTYKSGRNKALKKFKKEYLTMEVL-E 85
DB 599 LLNWLKWTAIQTTEIKENYKMQDTSMMKKK-LKALEKEQREIRIPEADELNTQTGILVE 757

```

972 NTKESLS-----NLSVAM-----KVQVYAQLFSTGL-----NTTIDASKVBE--L 1010
 1734 NRNEKVSQHIKSAKLLIAQPLVQCLVTTFTFTGTFVDFSDLEKLENDIENMLKFVEKHL 1793
 1011 VSTADETIDLLPTLSEGLPIIATIIDGVSLGAIAKELSETNDPLLRQEEI-E-----AKIGI 1066
 1794 ESSDEDEKWD-----BESAQI EEVLRQGEEMLPHPEDNKKERL 1834
 1067 MAVNUTAASTAVTSAL-----GIASFSLIIVLPL-----AGISAGIPS 1105
 1835 QULLLHRYNKIKAIPIQORRMGOLASIRSLPDTYLVINIKLLCMBDDVLSLNVPE 1894
 1106 LVNNELIQQDKATKVIDYFKHSIAETEGATLLDDKLIIM---PQDDLVLSEIDFNNSI 1162
 1895 L--NTAIYEDFS-----FQDSLKNIHQDQLKGEQIAVHEKQPDVIL----- 1936
 1163 TLGKCEIWRAGSGG-----HTILT-----DDIDHFFSPSPITY---RKPMLSIYDVNLINIK 1209
 1937 -----EASGPEAIQIRDTILTQNAKWDRINRMSYDRKCFDRAEWEWRQFHCDDLNDL 1988
 1210 KEKIDFSKDLMYLPNAPN-----RVFGYEM-----GWTPGFRSLDNDGTLLDRIRD 1256
 1999 TQWITEAEELLVDTCAPGSLDLEKARHQOELEVGISGHQPSAALNRTGDGIVQKLS 2047
 1257 HYEGOFYWRYPAFIA---DALITKLKPRYEDTVNRLDGNTRSFIVPVITTEQIRKNLS 1313
 2048 QADGSLKLEKLAGLNQRDAVAYKDR-----QPLKGBSK-----QVMK--- 2088
 1314 YSFYGGSGSYSLSPSYNNIDNLVENDTWIDVDNVV-KNITIE-SDEIQKGELIENI 1371
 2089 -----YHQD-----EIIICWLTKRAHAMQKRSITELGENLOE---LRDL 2125
 1372 LSKLMIEDNKI-ILNHNTHINFGD-----INESNRFTISITFS-ILEDI-----NI 1414
 2126 TQSEMEVHAEKWKWARTLEMLSDLSLPERDKTISELSRTVNMWNKICREVPETTLKEC 2185
 1415 IIEIDLVSXKYLILSGNCK--LIENSSDI 1443
 2186 IQEPSSVSQT-RIAAHPNVQKVLVSSASDI 2215
 RESULT 15
 MLP1_YEAST
 ID MLP1_YEAST STANDARD; PRT; 1875 AA.
 AC Q02455;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin-like protein MLP1.
 GN MLP1 OR YKR095W OR YKR415.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c;
 RC MEDLINE=93247549; PubMed=8483450;
 RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;
 RT "A new yeast gene with a myosin-like heptad repeat structure.";
 RL Mol. Gen. Genet. 237:359-369(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94205265; PubMed=8154186;
 RA Bou G., Eschbar P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
 RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
 RT "The complete sequence of a 15,820 bp segment of Saccharomyces
 RT cerevisiae chromosome XI contains the UB12 and MPL1 genes and three
 RT new open reading frames.";
 RT Yeast 9:1349-1354(1993).
 CC -!- FUNCTION: Myosin-like protein that is probably involved in DNA
 CC repair.
 CC -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.

CC -!- CAUTION: Ref.2 misquotes the gene name as "MPL1".
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L01992; AAA34783.1; .
 CC EMBL: X73541; CAAS1948.1; .
 CC EMBL: Z28320; CA882174.1; .
 CC PIR: S38173; S38173.
 CC GeneOnline: 140074; .
 CC
 CC DR GO: 00005635; C:nuclear membrane; IDA.
 CC DR GO: 00005654; C:nucleoplasm; IDA.
 CC DR GO: 00006606; P:protein-nucleus import; IDA.
 CC KW Coiled coil; DNA repair.
 CC FT DOMAIN 69 487 COILED COIL (POTENTIAL).
 CC FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
 CC FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
 CC FT CONFLICT 301 301 R -> A (IN REF. 1).
 CC SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C9066867 CRC64;
 Query Match 3.3%; Score 288; DB 1; Length 1875;
 Best Local Similarity 18.4%; Pred. No. 0.00097;
 Matches 314; Conservative 304; Mismatches 612; Indels 474; Gaps 75;
 QY 226 EESLNKNTANNNGNDIRNLEKFADEDLVRLYNOLV-----ERNWLAASDILRISMLK 278
 DB 18 DERLNAIAFFGCSIQVAKSF-DGDVVHKLNDKLQFNLKSENKLVTVSPFELKASSLK 76
 QY 279 EDGGVYLDVLDIPGIQPDLFKSN---KPSDITNTSWEMI-----KLEAIMKYKEY 326
 DB 77 KIDGLKTEENVIRENDKIRKERNDTFFVKFSEVENEKMKLSSELEFVVKRKLDDLTEEKKE 136
 QY 327 IPGYTSKFNMDLDEEVQ-----RSFESALSKSKSKSEIFLPD-----D 365
 DB 137 TQSNQORTLKILDERLKEIYELVRVNNRSNCKLRSTIMDLTKQQQYITNDLSRTE 196
 QY 366 IKVSPLEVKIAFANNVINOALISLKDSYCS-----DIVINOIKRYKIL----- 410
 DB 197 LERKTOELTLOSNDNDWLEKELRSKNEQYLSYRQTKDVILDIRNELNRLNDFQVETRN 256
 QY 411 NDNLNPSINEGDFNTTMTKIFSDKLASIS-----NEDNMFMWIKITNYLKVGFAPDVR 464
 DB 257 NDVLKORNE-----LSKSLQEKLEIKGLSDSLNSEKQEFSAEMS--LKQRLVDLLES 308
 QY 465 TINLSGPGVYTGAYQDMLFMKDNSTNIHLE-----PE-----LRNFEPFKTIQSLTEQ 514
 DB 309 QLN-----AVKEELNSIRELNTAKVIADDSKQTPENEDLLKELQTLKEKLAQ-CEK 359
 QY 515 EITSLWSFNQARAKSOFEEYKGYFEGALGEDNDLDFQNTVLDKDYVSKKILSKMKTEN 574
 DB 360 ECLRLSSITDE-----ADEDNENLSAKSSSDFIPLKKIUKERT-- 399
 QY 575 KEYIH-----YIVQLQGDKI-----SYEASCNLSKDPYSSILYQKNIEGSEYAYYYVAD 625
 DB 400 KEHLQNLQIETTFIVELE-HKVPINSPKRTDMLENLNAALLLETSNEK-----N 450
 QY 626 AEIKEDIKVRIPIYQISGNENIKLTFIGHGKSBFNTDTFANLDV-----DSLSS-- 673
 DB 451 AKVKEL-----NAKQKLV-----ECENDLQITLTKQRLDCRQIQYLLITNSVNSD 497
 QY 674 -----EIEITLNLAKADISPKYIEINLLGNMFSYSASSETYPGKLLKIKRVSLE 726
 DB 498 KGPLRKEEIQFQNIQMDEDDSTI-----TESDSQKVVTRELVEFKNIQLIQEKNAEL 549
 QY 727 MPISQSDSITVSANQYEVINEGKEIILDHGSKWINKESIKDISKEYISFNPKENK 766
 DB 550 LKVVNR-----LADKLESK-EKKSQSLQKISSETVNEAKEIITLKSRM-----DLER 599

Search completed: April 1, 2004, 16:42:35
Job time : 35.5 secs

QY 787 IIVKSYLHELSTLLQEIIRNANSSDIDLEKKVMTLCEINVASNIDROIVEGRIBAKN 846
Db 600 IEBLQELBELTSVP--NEDASYSNVTIKQ---LFTETKRDLESQV--ODLQTRISQITR 652
QY 847 L7SDSINYIKNEFKLIESISDLSYDLKHQNGLDSDHSFISFE-----DISKTENG- 895
Db 653 ESTENMSLLNKE---LQDLYDSKSDISIKLGEKSSRIILAEERFKLLSNTLDLTKAENDQ 709
QY 896 FRIRFINKGTGNSIFTETEK--EIFSEYATHISK-----EISNKDITFDNVNKGVLVK 947
Db 710 LRKRFDYLO--NTILKQDSKTHETLNEYVYCKSLISIVETELLNLKE-----EQKLAVH 761
QY 948 VNLDAAHEVNTLNSAFFIQLSIEIYNTKESLNSLVAMK-----VQVYAQLFSTGINTIT 1002
Db 762 LEKNLQELNKLSP-----EKDSLIMVTQLQKEREDELEETRSKQCKKIDLE 813
QY 1003 DASKVVELVSTALDETI-----DLLPTLSEGLPIIATIIDG----- 1038
Db 814 DALSELKETSQKDHKIQLEEDNNSNIWYQNKIEALKDYESVITSVDSKQTDIEKIQ 873
QY 1039 ---VSLGAAIKE-----LSET--NDPLLRQETIEAKIGIMAVNLTAATAI----- 1078
Db 874 YKVKSLKEIEEDKIRLHTVNWDEITINDSLRKELEKS-----KINLTDAYSQIKEYKDL 929
QY 1079 --VTSALGIASGPSILLVPLAGISAGIPSLVNNELLQDKATKVIDYFKHISLAETEGAF 1136
Db 930 YETTSQ-----SLQOTNSKLDSEFKDF---TNQIK 956
QY 1137 TLDDDKIIMPDDLVISEIDFN--NNSITLCKEIMWAEAGSGHGLTDDIDHFFSSPSITY 1195
Db 957 NLTEKTSLEDKISLLKEQNFNLNNELDIQ----- 987
QY 1196 RKPWLSIYDVLNIIKKKIDIPSKDMLVLPNAPNRFVGYEMGTFPGFSLDN----- 1245
Db 988 -----KMEKEKAFKXISILQNNKEVEAVKSEYSEKSLKIQNDLDQOTIYAN 1037
QY 1246 -----DGTCLDLRIDH---YEGQFYWYFAPFADALITKLKPYED--- 1284
Db 1038 TAQNNYEQELQHEADVSKTISELREQLHTYKGV--KTLNLSRDQLENALKENERSWSQ 1095
QY 1285 -----TNVRI--NLGNTRSPFVITTEIQRKNLSYSPYSGGYSLSLSPYNN 1333
Db 1096 KESILLEQLDLSNGRIEDLSQNKLYDQIYTAADKEYNNSTNGFLNNILITLRRERD 1155
QY 1334 I-----DLNLVENDTWJVDNVNKNITIESDEIQK--GELI---ENILSKLN- 1376
Db 1156 ILDTKVTVAERDAMLRQKISLMDVQLQARTKLDNSRVERKENHSSIIQOHDIMEKLNQ 1215
QY 1377 ---IEDNKIILNHTINFYGDINESNRFSLSILEDINIIBIDLVSKYKILLSGNC 1433
Db 1216 LNLRESNITLRNLEN-----NNKKELQSELDKLNQNVAPIESELTALKYSMOEKEQE 1271
QY 1434 MKLI-----ENSSDIQKIDHIGFNGEHQKIPYYSIDN-----ETK 1470
Db 1272 LKLAKEBVHWKRSQDILEKHQLS--SSDYKL--ESBIENLKELENKERQGAABEK 1328
QY 1471 YNGFIDYSKEGLFTAEFNSIESIRNI--YMPDSNNLFIYSSKDKDIRIINKGDVKLLIG 1529
Db 1329 FNR--LRQAQERLKTSLQCSLSTEQVNSLRDARNV--LENSLSEANARIEELQNAKVAQG 1386
QY 1530 N-----YFKDDMKYS--LSFTIEDTNT---IKLNGVYLDENGVAQJILKFWNAKSAL 1576
Db 1387 NNQLEAIRKLOEABKASRELOAKLEBESTTSYESTINGLNEEITTTKEEIEKQRIQOOL 1446
QY 1577 -----NTNSLNMFLESI-----NIKNIFFYNLDPNPIEFLDTNFIISGNSIQOPEL 1624
Db 1447 QATSANEQNDLSNIVESMKSPREEDKIKFKIKEXTQEVN--EKILEAQERLNPQSN--NMEE 1504
QY 1625 ICDKKNIQPIYFNFKIKETSYTL 1648
Db 1505 IKKKWESEHEQVSKIREABEAL 1528

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 1, 2004, 16:38:25 ; Search time 23.75 Seconds
(without alignments)
6895.288 Million cell updates/sec

Title: US-09-126-816C-6_COPY_1_1700

Perfect score: 8675

Sequence: 1 MNLVNAQLQKVVYKFRIO.....VLYGIDRYVKNKVIAPNLYT 1700

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8675	100.0	2364	2 I40884	cytotoxin L - Clos
2	6759.5	77.9	2366	2 S10317	toxin B - Clostrid
3	6737	77.7	2367	2 S70172	toxin A - Clostrid
4	4467	51.5	2710	2 A37052	alpha-toxin - Clos
5	2120.5	24.4	2178	2 S5805	toxin B - Escheric
6	507.5	5.9	3169	2 T00296	rhostry protein -
7	468.5	5.4	2401	2 T28676	rhostry protein -
8	440	5.1	2269	2 T28677	hypothetical prote
9	431.5	5.0	4688	2 F62885	hypothetical prote
10	422	4.9	2166	2 G70163	hypothetical prote
11	399	4.6	3724	2 T18427	hypothetical prote
12	398	4.6	1979	2 C71622	hypothetical prote
13	391	4.5	3225	2 D81702	adherence factor T
14	382.5	4.4	3394	2 T18501	hypothetical prote
15	368	4.2	1711	2 T18429	hypothetical prote
16	361.5	4.2	1939	2 T18372	repeat organelle
17	360	4.1	2829	2 A42771	reticulocyte-bind
18	348	4.0	1127	2 T28317	ORF MSV156 hypothe
19	345.5	4.0	4981	2 T18489	hypothetical prote
20	344	4.0	2819	2 A90551	conserved hypothet
21	343	4.0	2485	1 H71621	serine/threonine-s
22	340.5	3.9	5005	2 F62884	hypothetical prote
23	338.5	3.9	3335	2 H81702	adherence factor T
24	331	3.8	3255	2 G81702	adherence factor T
25	329.5	3.8	4550	2 T18440	hypothetical prote
26	329	3.8	2910	2 T28156	DNA-directed RNA p
27	323	3.7	1819	2 D97033	uncharacterized pr
28	318.5	3.7	3973	2 B71612	hypothetical prote
29	317.5	3.7	2599	2 F90608	ABC transporter pe

30 315 3.6 1302 1 JC6009 surface-located me
31 313.5 3.6 3844 2 T18402 asparagine/asparta
32 312 3.6 1639 2 S05603 major merozoite su
33 311.5 3.6 1726 1 SAZQGM major merozoite su
34 311.5 3.6 1726 2 A45948 major merozoite su
35 311 3.6 1516 2 E71619 RAD2 endonuclease
36 310 3.6 1306 2 T28313 ORF MSV152 probabl
37 309.5 3.6 1447 2 F82909 hypothetical prote
38 309.5 3.6 1701 2 A28668 major merozoite su
39 309.5 3.6 1957 2 T38077 hypothetical coile
40 309 3.6 2244 2 F90563 hypothetical prote
41 307.5 3.5 1701 2 A54498 major merozoite su
42 307.5 3.5 1830 2 E82909 conserved hypothet
43 305.5 3.5 2496 2 A71616 secreted protein p
44 304.5 3.5 1802 2 G71616 hypothetical prote
45 302.5 3.5 1252 2 B42771 reticulocyte-Bindi

ALIGNMENTS

RESULT 1

I40884

cytotoxin L - Clostridium sordellii

C;Species: Clostridium sordellii

C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 15-Oct-1999

C;Accession: I40884

R;Green, G.A.; Schue, V.; Montell, H.

Gene 161, 57-61, 1995

A;Title: Cloning and characterization of the cytotoxin L-encoding gene of Clostridium

A;Reference number: I40884; MUID:95369733; PMID:7642137

A;Accession: I40884

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-2364 <RES>

A;Cross-references: EMBL:X82638; NID:g1000694; PIDN:CAA57959.1; PID:g1000695

C;Superfamily: cpl repeat homology

C;Keywords: cytotoxin

Query Match 100.0%; Score 8675; DB 2; Length 2364;
Best Local Similarity 100.0%; Pred. No. 2.4e-304; Indels 0; Gaps 0;
Matches 1700; Conservative 0; Mismatches 0;

QY	1	MNLVNAQLQKVVYKFRIOEDEYVAILNALAEYHNMSSESVVEKYLKLDINLNDYVL	60
DB	1	MNLVNAQLQKVVYKFRIOEDEYVAILNALAEYHNMSSESVVEKYLKLDINLNDYVL	60
QY	61	NTYKSGRNKALKKPKFYLTMVELEKNNSLTPVEKNLHFIWIGQINDTAINYINQWKD	120
DB	61	NTYKSGRNKALKKPKFYLTMVELEKNNSLTPVEKNLHFIWIGQINDTAINYINQWKD	120
QY	121	VNSDYTVKVFYDSNAPLINTLKKTIIVESATNTLSEFRNLNDPFDYKNFKRMEIYY	180
DB	121	VNSDYTVKVFYDSNAPLINTLKKTIIVESATNTLSEFRNLNDPFDYKNFKRMEIYY	180
QY	181	DKOKHFDYKSGIENPEFIIDNIKTLSNEYSKDLEALNKYTEESINKITANNNDI	240
DB	181	DKOKHFDYKSGIENPEFIIDNIKTLSNEYSKDLEALNKYTEESINKITANNNDI	240
QY	241	RNLEKFADEDLVRLVNOELVERWNLAAASDIIRISMLKEDGGVYLDVDILPGIQDLPFKS	300
DB	241	RNLEKFADEDLVRLVNOELVERWNLAAASDIIRISMLKEDGGVYLDVDILPGIQDLPFKS	300
QY	301	INKPDSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQSFESALSSKSKSEIF	360
DB	301	INKPDSITNTSWEMIKLEAIMKYKEYIPGYTSKNFMDLDEEVQSFESALSSKSKSEIF	360
QY	361	LPDDIKVSPLEVKIAFANNVINQALISLKSYSDDLVIINQIKRYKILNDNLNPSINE	420
DB	361	LPDDIKVSPLEVKIAFANNVINQALISLKSYSDDLVIINQIKRYKILNDNLNPSINE	420
QY	421	GTDFNTMTKIPFSDKLASINEDNMFMKIITNLYKVGAPDVRSTINLSGPGVYTCAYOD	480

Db 421 GTDFNTMK:FSKLASISNEDNMFMKIITNYLKVGFAPDVRSTINLSGPGVYTCAQD 480
QY 481 LLAFKONSNIHLEPELRNFPKTKISOLTEQETISLWSNQBARAKSOFEEYKKGYPE 540
Db 481 LLAFKONSNIHLEPELRNFPKTKISOLTEQETISLWSNQBARAKSOFEEYKKGYPE 540
QY 541 GALGEDNDLFAQNTVLDKDYVSKKILSSMKTRNKEYIHYIYVLOQDKISYEASCNLFPSK 600
Db 541 GALGEDNDLFAQNTVLDKDYVSKKILSSMKTRNKEYIHYIYVLOQDKISYEASCNLFPSK 600
QY 601 DPYSSILYQKNIEGSETAYYYVADAEIKEIDKYRIPYQISNKRNIKLTFFIGHGSEFNT 660
Db 601 DPYSSILYQKNIEGSETAYYYVADAEIKEIDKYRIPYQISNKRNIKLTFFIGHGSEFNT 660
QY 661 DTFANLDVDSLSSEIETILNLAADISPKYIRINILGCNMFSYSABETYPCKLLLIK 720
Db 661 DTFANLDVDSLSSEIETILNLAADISPKYIRINILGCNMFSYSABETYPCKLLLIK 720
QY 721 DRYSELMPISQDSITVSANQVEVRINEGKEIILDSKWKINKESIIKDISKEYISF 780
Db 721 DRYSELMPISQDSITVSANQVEVRINEGKEIILDSKWKINKESIIKDISKEYISF 780
QY 781 NPENKIIVKSKYLHELSTLLQEIIRNANSDDILEKKVMLTECEINVASNIDRQIVEGR 840
Db 781 NPENKIIVKSKYLHELSTLLQEIIRNANSDDILEKKVMLTECEINVASNIDRQIVEGR 840
QY 841 IBEAKNLTSDSINYIKNEPKLESISDSLYDLKHQGLDDSHFISFEDISKTENGFRIRF 900
Db 841 IBEAKNLTSDSINYIKNEPKLESISDSLYDLKHQGLDDSHFISFEDISKTENGFRIRF 900
QY 901 INKETGNSFIETEKIPEYATHISKEISNKTIFDNVNGKLVKYNLDAAEHVNTLN 960
Db 901 INKETGNSFIETEKIPEYATHISKEISNKTIFDNVNGKLVKYNLDAAEHVNTLN 960
QY 961 SAFFIQSLLEYNTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELYSTALDETID 1020
Db 961 SAFFIQSLLEYNTKESLSNLSVAMKVQVYAQLFSTGLNTITDASKVVELYSTALDETID 1020
QY 1021 LLPTLSEGLPIIATIDGVSIGAAIKELSETNDPLLROEIEBAKIGIMAVNLTAATAIYT 1080
Db 1021 LLPTLSEGLPIIATIDGVSIGAAIKELSETNDPLLROEIEBAKIGIMAVNLTAATAIYT 1080
QY 1081 SALGIASGSIILVPLAGISAGIPSLVNNELLQDKATKVDYFKHISLAETEGAFTLLD 1140
Db 1081 SALGIASGSIILVPLAGISAGIPSLVNNELLQDKATKVDYFKHISLAETEGAFTLLD 1140
QY 1141 DKIIIMPQDDLVLSEIDFNNSNITLCKEIBWRAEGSGHTLTDHDFPSSPSIYIRKPEWL 1200
Db 1141 DKIIIMPQDDLVLSEIDFNNSNITLCKEIBWRAEGSGHTLTDHDFPSSPSIYIRKPEWL 1200
QY 1201 SIYDVLNITKEKIDFSDKLMVLPNAPNRFVGYEMGTGFRSLNDGDKLLDRIDHYEG 1260
Db 1201 SIYDVLNITKEKIDFSDKLMVLPNAPNRFVGYEMGTGFRSLNDGDKLLDRIDHYEG 1260
QY 1261 QYWRYPFIADALITKLPKRYEDTNVRINLDGNTRSPFIVPITTEQIRKNLSYSFYGSG 1320
Db 1261 QYWRYPFIADALITKLPKRYEDTNVRINLDGNTRSPFIVPITTEQIRKNLSYSFYGSG 1320
QY 1321 GSYLSLSFPYNNIDNLVENDTWVIDVDNVVKNITTESDEIQGELIENILSKLNTEDN 1380
Db 1321 GSYLSLSFPYNNIDNLVENDTWVIDVDNVVKNITTESDEIQGELIENILSKLNTEDN 1380
QY 1381 KIILNNHTINFYGDINESNRISLTFSTLEIDNIIIEIDLVSQYKILLSGNCMKLIENS 1440
Db 1381 KIILNNHTINFYGDINESNRISLTFSTLEIDNIIIEIDLVSQYKILLSGNCMKLIENS 1440
QY 1441 SDIOQKIDHIFGNGEHQYIIPYSYIDNETKNGFYDYSKKEGLFTAFPSNESIIRNYMP 1500
Db 1441 SDIOQKIDHIFGNGEHQYIIPYSYIDNETKNGFYDYSKKEGLFTAFPSNESIIRNYMP 1500
QY 1501 DSNLFIYSSKDLKDIRIINKGDVKLLIGNYFKDDMKVSLFTTEDNTIKLNGVILDEN 1560
Db 1501 DSNLFIYSSKDLKDIRIINKGDVKLLIGNYFKDDMKVSLFTTEDNTIKLNGVILDEN 1560

RESULT 2

S10317

toxin B - Clostridium difficile

C:Species: Clostridium difficile

C:Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text_change 15-Oct-1999

C:Accession: S10317; S21894; S22434

R:Barroso, L.A.; Wang, S.Z.; Phelps, C.J.; Johnson, J.L.; Wilkins, T.D.

Nucleic Acids Res. 18, 4004, 1990

A:Title: Nucleotide sequence of Clostridium difficile toxin B gene.

A:Reference number: S10317; MUID:90326540; PMID:2374729

A:Accession: S10317

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-2366 <BAR>

R:Michel-Streiber, C.; Laufenberg-Feldmann, R.; Sartingen, S.; Schulze, J.; Sauerborn,

submitted to the EMBL Data Library, July 1991

A:Description: Comparative analysis of Clostridium difficile toxins A and B.

A:Reference number: S21894

A:Accession: S21894

A:Molecule type: DNA

A:Residues: 1271-2366 <EIC>

A:Cross-references: EMBL:X60984; NID:G40445; PIDN:CAA43299.1; PID:G40446

Mol. Gen. Genet. 233, 260-268, 1992

A:Title: Comparative sequence analysis of the Clostridium difficile toxins A and B.

A:Reference number: S22434; MUID:92293124; PMID:1603068

A:Accession: S22434

A:Molecule type: DNA

A:Residues: 1791-2366 <VON>

A:Cross-references: EMBL:X60984

C:Genetics:

A:Gene: toxB

C:Superfamily: cpl repeat homology

C:Keywords: cytotoxin

Query Match 77.9%; Score 6759.5; DB 2; Length 2366;

Best Local Similarity 76.6%; Pred. No. 1.9e-235;

Matches 1302; Conservative 191; Mismatches 206; Indels 1; Gaps 1;

QY 1 MNLVNAQLOKVMYVKFRIOEDYVAILNALUEYHNMSSESVVEKYLKDKINNTDNYL 60

Db 1 MSLVNAQLOKVMYVKFRIOEDYVAILDALUEYHNMSSESVVEKYLKDKINNTDNYL 60

QY 61 NTYKSGRNKALKPKFYLTVMEVLKKNLSLTPVEKNLHFVWIGQINDTAINYNQWKD 120

Db 61 NTYKSGRNKALKPKFYLTVMEVLKKNLSLTPVEKNLHFVWIGQINDTAINYNQWKD 120

QY 121 VNSDYTVKYPYDSNAFLINTLKTIVESATNTLSPRENLDNPDYKFKRMEIY 180

Db 121 VNSDYTVKYPYDSNAFLINTLKTIVESATNTLSPRENLDNPDYKFKRMEIY 180

QY 181 DKQKFIIDYKQIENPEFIIDNIIKTVLSNEYSKOLEALNKYLEESINKITANGNDI 240

Db 181 DKQKFIIDYKQIENPEFIIDNIIKTVLSNEYSKOLEALNKYLEESINKITANGNDI 240

QY 241 RNLEKFADEDLVRLYNQELVERWNLAAASDIIRISMLKXEDGGVLDVDPICIQDLPFS 300

Db 241 RNLEKFADEDLVRLYNQELVERWNLAAASDIIRISMLKXEDGGVLDVDPICIQDLPFS 300

QY 301 INKPDSTWTSWEMIKLEAIMKYKEVPCYTSKCNFMDLDEEVRQSFESALASKSDKSEIF 360
DB 301 IEKPSVTVDFWEMTKLEAIMKYKEVPEYTSSEHFDMLDEEVCQSFESVLSKSDKSEIF 360
QY 361 LPLEDDIKVPLEKIAFANNVINQALISLKDYSGLVINOIKRKYKILANDLNPSINE 420
DB 361 SSGDMEASPLEKIAFNSKGIINQGLISVKDYSCLNIVKJENRYKILANSLNPAISE 420
QY 421 GTDFNTMTKIFSKLASISNEDNMMPMIKTYLVKVGAPADVRSTINISGPGVYTGAYQD 480
DB 421 DNDFTNTTTFIDSIMAEANADNGRFEMELGKYLVRGVFPDPVKTITNLSPPEAYAAAYQD 480
QY 481 LLMFKDNSTNHLLEPELNFPEPKTKISQLEOBTISLWSNQARAKSOPEEYKKGFFE 540
DB 481 LLMFKGSMNHLLEADLNFPEISKTNISQSTOEAMSLWSDDDARAKAFEEYKGNFFE 540
QY 541 GALGEDNDLFAONTVLDKDYYSKTLSSMKTRNKEYIHVIVOLQDKISYEASCNLPSK 600
DB 541 GSGEDNDLDFQNVVDVDEKLEKISLARSEGYIHVIVOLQDKISYEAACNLPAK 600
QY 601 DPYSSILYQKNIEGSETAYVYVADAEIKEDKIRIPYQISNKRNIKLTFFIGHKSEFNT 660
DB 601 TPVDSVLQKNIEDSEIAYVYVADGEIOEDKIPSIISDRPKILTFIGHKDEFT 660
QY 661 DTFANLDVLSSEIETILNLAKADISPKYIEINLLGCMNFSYSISABETYPGKLLLIK 720
DB 661 DIFAGFDVLSLSEIEAADLAKEDISPKSIEINLLGCMNFSYSINVEETYPGKLLIKV 720
QY 721 DRYSELMPISQDSITVSANQEVNRINEGKEILDHSGKWNKEEIIKDISSEYISF 780
DB 721 DKISELMPISQDSITVSANQEVNRINEGKEILDHSGKWNKEEIIKDISSEYISF 780
QY 781 NPENKIIVKSKYLHELSTLQBIIRNANSDDILEKKWMLTECEINVASNTDROIVER 840
DB 781 NPENKIIVKSKNLPESLTLQBIIRNANSDDILEKKWMLTECEINVASNTDROIVER 840
QY 841 IEEAKNLTSDSINYIKNEPKLIBESIDSLYDLKHQGLDSDHFSPEIDSKTENGFRIF 900
DB 841 IEEAKNLTSDSINYIKNEPKLIBESIDSLYDLKHQGLDSDHFSPEIDSKTENGFRIF 900
QY 901 INKETGNSPIETKEIPSEYATHISKEISNIKDTTFDNNVGLVKVKNLDAHEVNTLN 960
DB 901 INKETGESIFVETEKTFIFSEYANHITEISKIKGTTFDVTNGLVKKNLDTTHEVNTLN 960
QY 961 SAFIQLIEYNTTKESLSNLSVAMKVQVYQAFSTGLNTITDASKVVELVSTALDETID 1020
DB 961 AAFFIQSLIEYNSKESLSNLSVAMKVQVYQAFSTGLNTITDASKVVELVSTALDETID 1020
QY 1021 LLPTLSEGLPIIATIDGVSIGAAIKELSETNDPLLRQBIETAKIGIMAVNLTAASAIYT 1080
DB 1021 LLPTLSEGLPIIATIDGVSIGAAIKELSETNDPLLRQBIETAKIGIMAVNLTAATIT 1080
QY 1081 SALGIASGFSILLVPLAGISAGISPLVNNELILQDKATKVIDYFKHISIAETEGAFITLD 1140
DB 1081 SSGIASGFSILLVPLAGISAGISPLVNNELVLRDKATKVDFYFKHVSIVETEGVFTLD 1140
QY 1141 DKIMPODDVLVSEIDPNNNSITLKGCEIWRAGGSGHITLDDIDHFFSPSITTKPWL 1200
DB 1141 DKIMPODDVLVSEIDPNNNSITLKGCEIWRMEGGSGHTVTDDIDHFFSAPSITTYREPL 1200
QY 1201 SIYDVLNKKKEKIDFSKDLAVLPNAPRVFGYEMGTTPGRSLDNDGTGLDRIDRHYEG 1260
DB 1201 SIYDVLVEQKEEDLSKDLAVLPNAPRVFAWETGTPGRSLDNDGTGLDRIDRNYEG 1260
QY 1261 QFYWRVFAFADALITLTKPRYEDTNVRINLNDGNTRSFIVFVITTEQIRKNSYSFYSGS 1320
DB 1261 EFYWRVFAFADALITLTKPRYEDTNVRINLNDGNTRSFIVFVITTEYREKLSYSFYSGS 1320
QY 1321 GSYLSLSGPNMNDLNVENDTWIDVNVKNNTIESDEIOKGLIENILSKUMNED 1380
DB 1321 GTVALSLSQNMGINELSESDWIIDVNVVRDVTTIESDKIKKGLDIEGILSTLSEEN 1380

QY 1381 KIILNHTINFYGDINESNRFLSTFSLIEDINIIIEIDLVSYSYKILLSGNCKMLIENS 1440
DB 1381 KIILNSHEINFSEVNGSGFVSLTFSILLEGAINAIEVDLLSKYSKILLISGELKILMLNS 1440
QY 1441 SDIQKIDHGFNGEHOQKIPYISYINETKYNQFIDYSKKEGLTBAFNFESIIIRIYMP 1500
DB 1441 NHIOQKIDYGFNSELQKNIPYSFVDSSEGKNGFINGSTKEGLFVSELFDVVLISKVYMD 1500
QY 1501 DSNLFTYSSKDLKDIRIINKGDKVLLIGNYFKDKMKVSI-SFTIEDNTIKNGVYLDEN 1560
DB 1501 DSKPSFGYYSNNLKDVKVITKDNVNITLGYLKKDDIKISLSLTLQDEKTIKLSNVHDES 1560
QY 1561 GVAQILFKMNAKSALNTSNLMNFLESINIKNIFNNLDPNIEFLDNTFIISGNSIG 1620
DB 1561 GVAEILKPMNR-KONTNWSLSMFLESNMKISIFVNFLOSNIKFLDANFIISGTSIG 1619
QY 1621 QFELICDKDKNIOPYFINFKIKETSYLVGNRQNLIVEPSYHLDPSGNSISSTVINFSQK 1680
DB 1620 QFEPICDENDNIOPYFIKFTNTLETNTVLYVGNQNMIVEPNYDLDBSGDISSTVINFSQK 1679
QY 1681 YLYGIDRYVNVKVIAPNLYT 1700
DB 1680 YLYGIDSCVKNWISPNYIT 1699

RESULT 3
S70172
toxin B - Clostridium difficile
C:Species: Clostridium difficile
C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 15-Oct-1999
C:Accession: S70172; S4271
R: von Eichel-Streiber, C.; Meyer zu Heringdorf, D.; Habermann, E.; Sartingen, S.
Mol. Microbiol. 17, 313-321, 1995
A:Title: Closing in on the toxic domain through analysis of a variant Clostridium diff.
A:Reference number: S70172; MUID:9607928.; PMID:7494480
A:Accession: S70172
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-2367 <VON>
A:Cross-references: EMBL:Z23277; NID:g761713; PIDN:CAA80815.1; PID:g761714
A:Experimental source: isolate 1470
R:Sartingen, S.; von Eichel-Streiber, C.
submitted to the EMBL Data Library, July 1993
A:Description: Cloning and sequencing of an Clostridium difficile toxin B mutant.
C:Superfamily: cpl repeat homology
A:Reference number: S44271
A:Accession: S44271
A:Molecule type: DNA
A:Residues: 1-1323, 'N', 1325-2367 <SAR>
A:Cross-references: EMBL:Z23277
C:Keywords: cytotoxin

Query Match 77.7%; Score 6737; DB 2; Length 2367;
Best Local Similarity 76.1%; Pred. No. 1.2e-234;
Matches 1294; Conservative 201; Mismatches 204; Indels 2; Gaps 2;
QY 1 MNLVKAQLOQWVVKVRIQDEYVAILNALVEEYHNNSESSVVEKYLKLDKINLNTNYL 60
DB 1 MSLVNRKQLEKMANVRFRVQDEYVAILDALVEEYHNNSENVVEKYLKLDKINSLDTYI 60
QY 61 NTYKSGSKNALKKFKKEYLTMEVLEKXNSLTPYEKHLHFIIWIGQINDTAINYNQWKD 120
DB 61 DTYKSGSKNALKKFKKEYLVIEILEKXNSLTPYEKHLHFIIWIGQINDTAINYNQWKD 120
QY 121 VNSDYTVKVFVDSNAFLINTLTKTIVESATNTNLESFRENLDNDFDYNKFKYKRMELIY 180
DB 121 VNSDYTVKVFVDSNAFLINTLTKTIIESASNDTLESFRENLDNDFDYNKFKYKRMELIY 180
QY 181 DKQKHFIDYKSOJEENPEFIIDNIKTYSNYSKOLLEALNKYIEESLNKITANNNDI 240
DB 181 DKQKFNINYYKAEENPDLDIDIVKTYLSNEYSKOIDELNAYIEESLNKVTENSGNDV 240
QY 241 RNLEKFADEDLVRLYNQELVERMNLAAASDILRISMLKEDGGVYLDVDPILPGIQPDFKS 300

A:Molecule type: DNA		541	GALGEDNDLPAQNTVLDKDY-VSKILSS--MKTRNKYIHYIVOLQGDKISYEACNL	597
A:Residues: 1-55 'V', 57-2079 'L', 2081-2549 'S', 2551-2710 'SAU'		540	CSLSEGDVDFNKATLDKYNLNNKIPSNVNEBAGSKYVHYIIQOGDDISEATCNL	599
A:Cross-references: EMBL:X31797		598	FSKDPYSILYQKNIEGSETAYYYVAD--AEIKEIDKIRIPYOISNKRNIKLFIGHGK	655
C:Genetics:		600	FSKNPKSIIIQNN--NESAKSYFLSDGSESILELNKRIPIERLKNKEKVKVTFIGHGK	657
C:Superfamily: Clostridium difficile toxin A; cpl repeat homology		656	SEFNTDTPANLDVSLSESETIINLAKADISPXYEINLLGCMFYSYSABEITYFGKL	715
C:Keywords: cytotoxin; enterotoxin		658	DEFNTSEFARLSVDSLSEISSEFLDTIKDISPNVNEVLLGCMFYSYDFNVEEYFGKL	717
F:1820-1839/Domain: cpl repeat homology <CP01>		716	LLKIKDRVSELMPSISDSDITSVAQYEVRIINEREGREILDHSGKWINKESIIKDISSK	775
F:1840-1860/Domain: cpl repeat homology <CP02>		718	LLSMDKITSTLPDVKNNISITIGANQYEVRIINEREGREILDHSGKWINKESIIKDISSK	777
F:1861-1881/Domain: cpl repeat homology <CP03>		776	EYISFNPKNKIIVKSKYLHELSTLLOEIRNANSSDIDLEKVMTECEINVASNIDRQ	835
F:1882-1902/Domain: cpl repeat homology <CP04>		778	EYIFDSIDNKLKAKSKNIPGLASISEDIKTLILDASVSPDTKTLNKLNISSIGDY	837
F:1903-1923/Domain: cpl repeat homology <CP05>		836	IVEGRIEAKNLTSDSINYIKNEFKLIESDSYDLAKHQGLDSDHSPEDISKYENG	895
F:1924-1944/Domain: cpl repeat homology <CP06>		838	IYVEKLEPVKNIHNSIDDLIDEFNLENVSDLEYELKLNLDKYLISPEDISKXNST	897
F:1945-1965/Domain: cpl repeat homology <CP07>		896	FRIRFINKETGNSIFIEKEIFSEYATHISKEISNTKOTIFDNYNGKLVKKVNLDAHE	955
F:1966-1986/Domain: cpl repeat homology <CP08>		898	YSVRFINKSGESYVETEKEIFSKYSEHTKKEISTIKNSIITVDNGLNDIQLDHTSQ	957
F:1987-2007/Domain: cpl repeat homology <CP09>		956	VNTLSAPFIOSLIEYNTKESLSNLSVAMKVQVYAOQLFSTGLNTITDASKVSLVSTAL	1015
F:2008-2028/Domain: cpl repeat homology <CP10>		958	VNTLSAPFIOSLIDYSGNKVDLNDLSTSVKQVYAOQLFSTGLNTITDASKVSLVSTAL	1017
F:2029-2049/Domain: cpl repeat homology <CP11>		1016	DETIDLPFLSEGLPIIATIIDGVSLGAAIKELSETNDPLLRQIEAKIGIMAYNLTAAS	1075
F:2050-2070/Domain: cpl repeat homology <CP12>		1018	NDTINVLPTIEGPIVSTILDGINLGAAILKELADEHPLLKLEAKVGLAINMSLSI	1077
F:2071-2091/Domain: cpl repeat homology <CP13>		1076	TAIVTSALGASGSIILLVPLAGISAGIPSLVNNELIQQKATKVIDYFKHISAEATEGA	1135
F:2092-2112/Domain: cpl repeat homology <CP14>		1078	AATVASIVGIGAEVTIFLLPIAGISAGIPSLVNNELIQQKATKVIDYFKHISAEATEGA	1137
F:2113-2133/Domain: cpl repeat homology <CP15>		1136	FTLDDDDKIIMQDDLVLSLSEIDFNNSITLKGCEIWRABGGSGHTLTDIDHFPSSPSITY	1195
F:2134-2154/Domain: cpl repeat homology <CP16>		1138	LKTEDDKLVLPIDLVLSLSEIDFNNSITLKGCEIWRABGGSGHTLTDIDHFPSSPSITY	1197
F:2155-2175/Domain: cpl repeat homology <CP17>		1196	RKPLMSIYDVNLINKEKIDFSKDLWLPNAPNRVFGVEMGWTGPRSLDNDGTLLDRIR	1255
F:2176-2196/Domain: cpl repeat homology <CP18>		1198	HIPSLSIYSAIGIETENLDFSKIMLPNAPNRVFGVEMGWTGPRSLDNDGTLLDRIR	1257
F:2197-2217/Domain: cpl repeat homology <CP19>		1256	DHYEGQFYRYFAFIADALITKLPYEDTNRINLPCNTRSFIVPVITTEQIRKLSYS	1315
F:2218-2238/Domain: cpl repeat homology <CP20>		1258	DLYPGKFYRYFAFF-DYAITLKPVEDTNIKIKLKDITENFIMPITITTEQIRKLSYS	1316
F:2239-2259/Domain: cpl repeat homology <CP21>		1316	FYGGGYSLSLSPYNNIDNLVENDTWIVDNDVNVKNITIEDETOGELIENILSKL	1375
F:2260-2280/Domain: cpl repeat homology <CP22>		1317	FDGAGGYTSLLSYSPSTINLSKDLWIPNIDNEVREISIEGTIKKGLIKDLVLSKI	1376
F:2281-2301/Domain: cpl repeat homology <CP23>		1376	NIEENKILNHNTHFYDYNESNRFLSLFESILEDINILIEIDLVSYSKYLISGCMK	1435
F:2302-2322/Domain: cpl repeat homology <CP24>		1377	DINKKLIIGNOTIDFGSDIDNKORYIFLTCELDKDLISLIEILNLSVSKYLLSGKNY	1436
F:2323-2343/Domain: cpl repeat homology <CP25>		1436	LIENSDDIQQKIDHIGFNGEHQKYPISYID-NETKYNGFIDYSKGEGLFAEBSNESII	1494
F:2344-2364/Domain: cpl repeat homology <CP26>		1437	LISNLSTIEKINVLGLD---SKNIAYNTDESNNKYFGAI-----SKTSQKSI	1493
F:2365-2385/Domain: cpl repeat homology <CP27>		1495	RNIYMPDSNNL-----FYSSKOL--KDRIIRKGVKVLKIGNYFKO---DMKVSLS	1541
F:2386-2406/Domain: cpl repeat homology <CP28>		1494	H--YKDSKNTLEFVNDSTLEFNSKDFIADINVMKDDINTITGKYVDVNDNTDKSDFS	1541
F:2407-2427/Domain: cpl repeat homology <CP29>		1542	FTIEDNTIKLGVYLDENGVAQILKFNNAKSAINTSNLSMNPLESINIKNIYNNLDP	1601
F:2428-2448/Domain: cpl repeat homology <CP30>		1542	ISLVSKNQVNGLYLAESVYSYLDYFVNSDGHNTSNFMNLFIDNISFKWJGFGE---	1598
F:2449-2469/Domain: cpl repeat homology <CP31>		1602	NIEFLDTNFIISGNSITGQFELICDKDKNIQPIYFINKIKETSYTLVYGNRQNLIVEPS	1661

Query Match 51.5%; Score 4467; DB 2; Length 2710;	
Best Local Similarity 50.7%; Pred. No. 6 2e-153;	
Matches 87; Conservative 366; Mismatches 440; Indels 42; Gaps 15;	
QY	1 MNLVKAQLOKMYVVFRIQDEYVAILNLALEYHNMSESSVYKYLKLDINLNDYL 60
DB	1 MSLISKEELIKLAY-SIRPRENEKYTLTILNLDYVNLKLTNNENKYLQKLKLNESIDVFM 59
QY	61 NTYKSGRNKALKKFEYLTMVLELKNLSLTPVEKNLFIWIGGGINDTAINYNQWID 120
DB	60 NKYTSRNRNALNKLKIDILKELVILKNSNTSPVEKNLHFVWIGGVSDIALEYIKQWAD 119
QY	121 VNSDYTVKVFYDSNAFLINTLTKTIVESATNNLTSPRENLDNPEFDYKFKRMEITY 180
DB	120 INAEYNKILWYDSEAFVNLTKKALVESSTTEALQLEBEIQPQDNMKFYKGRMEITY 179
QY	181 DKQGHFDYKSOEENPEFTIINIITYLSNEYSKOLEALNKYIEESLNKIPANNNDI 240
DB	180 DRQKRFYNYKSOINKPTVFTIDILKSHLVSEYNRDETVLESYRTNSLRKNSHGIDI 239
QY	241 RNLKFADEDLVRLYNQELVERNLAAASDILRSMKEDGGYLDVILPGIQLPKFS 300
DB	240 RANSLFTEQELLNIYSQELLNRLGAAASDIVLLALKNGFYVLDVDMPLGHSDFLT 299
QY	301 INKPDSTINTSWEMIKLEATMKYKYPGVTSTKFNFDLDEEVORSFESALSSKSDSEIF 360
DB	300 ISRPSISGLDRWEMIKLEATMKYKYNNTSENFDKLDQDLNFKLIESKSESEIF 359
QY	361 LPDDDIKVSPELVKIAFANNNSNOALISLKDSYCDLVINQIKNRYKILNKNLNSINE 420
DB	360 SKUENLVSDLEIKIAFALGSVINOALISKQGSYLVNLVTEQVKNYQFNLQNLPAIES 419
QY	421 GTDFNTMTKIFSDKLASINEDNMFMKIKITNLKVGFPADPVRSTNLSPGVYTCAYQD 480
DB	420 DNNFTDTTKIFHDSLFNSATAENSMFTKTLAPYLVGFNPEARSTLSOPGAYASAYD 479
QY	481 LLMPKDNSTWHLPELRNPEFPKTKISQITQEITSLSMFSNQAARAKQFEBYKGYPE 540
DB	480 FINLQENTIEKTKASDLIEFKFPNNLSQITQEINLSLWSPQASAKYQFEBYKVRDYG 539